New tools for deciphering the roles of tryptophan *C*-mannosylation

Alan John

ORDCID ID: 0000-0002-6855-5470

A dissertation submitted in total fulfillment of the requirements of the degree of **Doctor of Philosophy**

2019

Supervisors: A/Prof Ethan Goddard-Borger and Prof Benjamin Kile

The Walter and Eliza Hall Institute of Medical Research
Department of Medical Biology
Faculty of Medicine, Dentistry and Health Sciences
The University of Melbourne

Abstract

Tryptophan *C*-mannosylation involves the covalent and irreversible attachment of α-D-mannose onto the C2 carbon of the indole ring of tryptophan forming an unusual *C*-glycosidic bond that is far more stable than an *N*- or *O*-glycosidic bond. *C*-mannosylation was first discovered on RNaseII in humans in 1994 and there are now over twenty human proteins for which there is evidence confirming the presence of this modification. It is found predominantly on thrombospondin type-1 repeat domains and type-I cytokine receptors. The glycosylation consensus sequence for *C*-mannosylation is WXXW.

Despite being conserved on many important proteins, the function and biological significance of *C*-mannosylation remains unknown. Addressing this problem was hampered by the fact that the genetic basis for *C*-mannosylation was unknown for a long time and because there are limited tools available for installing and detecting this protein modification. This has been remedied to some degree by the discovery that the *dpy19* gene in *C. elegans* encodes a tryptophan *C*-mannosyltransferase. There are four homologues of *dpy19* in humans and recent studies have found that at least two of these encode functional tryptophan *C*-mannosyltransferases.

I have embarked on the creation of a suite of tools for studying the roles of tryptophan mannosylation to progress this field. A simple *Pichia pastoris* protein production system was created to produce natively *C*-mannosylated proteins, with both C^2-α-D-mannosyl-tryptophan and without the modification. This was accomplished by integrating *C. elegans dpy19* under the strong constitutive GAPDH promoter. This yeast strain enabled the production of a wide variety of recombinant proteins with and without C^2-α-D-mannosyl-tryptophan. Two to five-fold improvements in protein production were observed when the protein was *C*-mannosylated, implying that the modification may play a role in protein folding, trafficking and/or stability. Differential scanning fluorimetry subsequently revealed that *C*-mannosylation of proteins stabilises them against thermal denaturation. This same yeast system provided a convenient means to explore the substrate specificities of the DPY19 enzyme and probe what enzyme residues were important for enzyme activity.

I have also generated the first monoclonal antibodies for detecting C^2-α-D-mannosyl-tryptophan. Using a *C*-mannosylated peptide derived from recombinant proteins produced in

my engineered yeast, I was able to generate five novel monoclonal antibodies that only recognised proteins with the modification. These monoclonal antibodies were validated for use in Western blots and ELISAs and biophysically characterised using surface plasmon resonance techniques. Sequencing the hybridomas that made these antibodies revealed a case of immunoconvergence, where two mice converged on essentially the same solutions for recognising C^2-α-D-mannosyl-tryptophan. I was able to crystallise a complex of one antibody's F_{ab} with the *C*-mannosylated peptide and determine its structure using X-ray diffraction techniques to reveal the key interactions used to recognise C^2-α-D-mannosyl-tryptophan.

Using the engineered yeast, monoclonal antibodies and substrate mimics of *C*-mannosylation, we were able to discover the first chemical inhibitors of *C*-mannosylation that can be used to probe *C*-mannosylation biology.

Lastly, I genetically disrupted the *dpy19* homologues in human cells and performed some comparative cell surface proteomic experiments to provide the first insights into how loss of *C*-mannosylation impacts the human proteome.

Declaration

This is to certify that:

(i) This thesis comprises only my original work towards my PhD degree except where indicated in the Preface,

(ii) Due acknowledgement has been made in the text to all other material used, and

(iii) The thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Alan John

Preface

Pursuant to the regulations governing the degree of Doctor of Philosophy at the University of Melbourne, I hereby submit that:

Chapter 2: this chapter represents 85% of my own work.

- Dr. Nichollas Scott performed the mass spectrometry analysis of all of the recombinant proteins produced with and without *C*-mannosylation

Chapter 3: this chapter represents 100% of my own work.

Chapter 4: this chapter represents 95% of my own work.

- Dr. Nichollas Scott performed the mass spectrometry analysis of IL12B in the three DPY19 mutants.

Chapter 5: this chapter represents 70% of my own work.

- Dr. Michael Jarva solved the crystal structure of the 5G12 F_{ab} in complex with the *C*-mannosylated peptide. This includes the digestion of 5G12 into Fab fragments and their purification. I acknowledge Michael and A/Prof Ethan Goddard-Borger for making Figure 5.8.
- Dr. Nichollas Scott performed mass spectrometry analysis on the purified *C*-mannosylated peptide.
- Dr. Richard Birkinshaw assisted with the SPR experiments
- A/Prof Chris Tonkin provided *Toxoplasma gondii* tachyzoites
- Dr. Raphael Trenker and A/Prof Matt Call assisted with the preparative HPLC.
- The WEHI monoclonal antibody facility assisted in creating the monoclonal antibodies.

Chapter 6: this chapter represents 85% of my own work.

- Ms. Hui Min Tay, Dr. Sayali Shah, and A/Prof Ethan Goddard-Borger designed and synthesised the DPY19 substrate mimics and inhibitors.
- Dr. Nichollas Scott performed mass spectrometry analysis on the two mutants of IL12B.

Chapter 7: this chapter represents 85% of my own work.

- Dr. Jarrod Sandow performed the mass spectrometry on the purified samples.

Publications

Published research articles

Shathili, A.M., Bandala-Sanchez, E., **John, A***., Goddard-Borger, E.D., Thaysen-Andersen, M., Everest-Dass, A.V., Adams, T. E., Harrison, L.C., and Packer, N.H. Specific Sialoforms Required for the Immune Suppressive Activity of Human Soluble CD52. Front. Immunol, 2019. *Second author

Khurana, S., Coffey, M.J., **John, A***., Uboldi, A.D., Huynh, M.H., Stewart, R.J., Carruthers, V.B., Tonkin, C.J., Goddard-Borger, E.D., Scott, N.E. Protein *O*-fucosyltransferase 2-mediated *O*-glycosylation of the adhesin MIC2 is dispensable for *Toxoplasma gondii* tachyzoite infection. J Biol Chem, 2018. *Second author

Abayakoon, P., Jin, Y., Lingford, J.P., Petricevic, M., **John, A**., Ryan, E., Wai-Ying Mui, J., Pires, D.E.V., Ascher, D.B., Davies, G.J., Goddard-Borger, E.D., Williams, S.J. Structural and Biochemical Insights into the Function and Evolution of Sulfoquinovosidases. ACS Cent Sci, 2018. 4(9): p. 1266-1273.

McKenzie, N.C., Scott, N.E., **John, A**., White, J.M., Goddard-Borger, E.D. Synthesis and use of 6,6,6-trifluoro-L-fucose to block core-fucosylation in hybridoma cell lines. Carbohydr Res, 2018. 465: p. 4-9.

Lopaticki, S., Yang, A.S.P., **John, A***., Scott, N.E., Lingford, J.P., O'Neill, M.T., Erickson, S.M., McKenzie, N.C., Jennison, C., Whitehead, L.W., Douglas, D.N., Kneteman, N.M., Goddard-Borger, E.D., Boddey, J.A. Protein *O*-fucosylation in *Plasmodium falciparum* ensures efficient infection of mosquito and vertebrate hosts. Nat Commun, 2017. 8(1): p. 561. *Second author

Jin, Y., Petricevic, M., **John, A**., Raich, L., Jenkins, H., Portela De Souza, L., Cuskin, F., Gilbert, H.J., Rovira, C., Goddard-Borger, E.D., Williams, S.J., Davies, G.J. A beta-Mannanase with a Lysozyme-like Fold and a Novel Molecular Catalytic Mechanism. ACS Cent Sci, 2016. 2(12): p. 896-903.

Roth, C., Petricevic, M., **John, A**., Goddard-Borger, E.D., Davies, G.J., Williams, S.J. Structural and mechanistic insights into a *Bacteroides vulgatus* retaining *N*-acetyl-beta-galactosaminidase that uses neighbouring group participation. Chem Commun (Camb), 2016. 52(74): p. 11096-9.

Roth, C., Petricevic, M., **John, A**., Goddard-Borger, E.D., Davies, G.J., Williams, S.J. Structural and mechanistic insights into a *Bacteroides vulgatus* retaining *N*-acetyl-[beta]-galactosaminidase that uses neighbouring group participation. Acta Crystallographica Section A, 2016. 72(a1): p. s221.

Primary research articles in preparation

John, A*., Scott, N.E., Jarva, M., Shah, S., Tay, H., Birkinshaw, R.W., Trenker, R., Call, M.E., Czabotar, P.E., Goddard-Borger, E.D. An engineered microbial system for deciphering the molecular functions of tryptophan *C*-mannosylation. (Manuscript under preparation)
*****Sole first author

John, A*., Scott, N.E., Goddard-Borger, E.D. Insight into conserved and crucial amino acids in the DPY19 family of *C*-mannosyltransferases. (Manuscript under preparation)
*****Sole first author

John, A*., Sandow, J., Webb, A.I., Goddard-Borger, E.D. Loss of *dpy19l1* and *dpy19l3* results in reduction of cell surface expression of target proteins. (Manuscript under preparation)
*****Sole first author

Review article in preparation

John, A., Goddard-Borger, E.D. Recent advances in tryptophan *C*-mannosylation. (Manuscript under preparation)

Acknowledgements

Firstly, I would like to thank and acknowledge my primary supervisor A/Prof Ethan Goddard-Borger for his support, guidance, and mentorship throughout my PhD studies. You have shown me support and friendship all throughout my PhD, not just the highs of science but most importantly the lows. Just as you have watched me grow into a scientist, I too have watched you become a great leader. I not only consider you as my boss and mentor, but I also consider you as a dear friend. Thank you for giving me such a cool project to work with, I still remember the thrill I got when you first told me about the project. I was very fortunate to have trained directly under your supervision both throughout my master's degree and my PhD degree. Not many students have the luxury of being taught directly by their lab head which has provided me with a wealth of knowledge and experience which I am ready to pass onto others.

To my co-supervisor Prof Benjamin Kile, thank you for your immense support, your contagious enthusiasm for science and great leadership. It has been a pleasure having you a part of my project. To my committee chair, A/Prof Isabelle Lucet, your support and genuine care for my well-being and success has been unprecedented.

It has been a privilege to work in the Goddard-Borger lab, the people in this lab really made it that much easier to conduct my PhD studies. Nicole McKenzie and I start our PhD studies together, and our friendship during that time has been invaluable. To James Lingford, my friend and the glue that keeps the lab intact, thank you for being such a great person to be around and for all your help. Gaetano Speciale, you have been a great friend during both my master's and my PhD. Michael Jarva, you have taught me a lot and have become a great friend and mentor. To Sayali Shah, your help and knowledge has been invaluable to my project and it has been a pleasure to have worked with you. Thank you to Hui Min Tay for your help and contribution to my project in the short time that you were in our lab, you a truly an exceptional student.

To the man who made such a great impact to my PhD project, Nichollas Scott. I would not have come this far without your proteomics. Thank you for all the hard work that you out put into my project. You are truly an exceptional and talented scientist. Your work ethic is something out of this world and is very contagious.

I would like to thank all of the members of the chemical biology division, both past and present, especially 5E. The chemical biology division is truly a lovely division to be a part of and to work in. 5E has made my study at WEHI a delight, your support and understanding is tremendously appreciated. To Guillaume Lessene, thank you for your support and interesting conversations and for the wonderful Jazz music that you play Friday evenings. Leeanne Lewis, you are such a wonderful person to be around and thank you for all your help. Stephane Chappaz, thank you for your conversations and wisdom. Kate McArthur, thank you for teaching me the world of mammalian cell culture and CRISPR, being around you has been inspirational.

To my mentors and friends, Kate Lawlor, thank you for your wisdom and knowledge. Dominic De Nardo, you have been a great mentor and friend and your scientific advice has been invaluable. Esther Bandala-Sanchez, you have such a calming demeanour which is contagious. Your friendship, mentorship and scientific advice has helped me immensely throughout my PhD. To Rachael Lane, thank you for always being there for me and supporting me, you are a truly great friend. Christine De Nardo, your energy is contagious, thank you for being there for me. Jessica Brewster, you have been a great friend to me, thank you for your support. Marija Dramicanin-Grinter, thank you for your friendship and guidance.

WEHI is a great place to undertake a PhD, the support and environment throughout the institute is unparalleled. I would like to thank the Flow Cytometry lab, mAb lab, and media staff for their support. Sue Hardy and Keely Bumsted O'Brien have been a great support throughout my PhD. I appreciate and thank the Australian Federal Government for funding my PhD with an Australian Postgraduate Award scholarship. I would also like to thank the University of Melbourne for facilitating my PhD studies.

Lastly, to my family, thank you for your support and unconditional love all throughout my PhD studies. My PhD would not have been possible without your support and care. I thank my family for always being there for me and understanding when I have had to reschedule dinners and meetups. To my mother, your love and support is what keeps me going and making you proud is what motivates me in life.

Abbreviations

-/-	Null allele
3D	Three dimensional
aa	Amino acid
Ab	Antibody
ACHE	Acetylcholinesterase
AD	Alzheimer's disease
ADAMTS	A disintegrin-like and metalloproteinase with thrombospondin type 1 repeats
ADGRB1	Adhesion G protein-coupled receptor B1
ADP	Adenosine diphosphate
ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
APC	Allophycocyanine
AQP	Aquaporin
ATP	Adenosine triphosphate
BAI1	Brain angiogenesis inhibitor 1
BCHE	Butyrylcholinesterase
BLAST	Basic Local Alignment Search Tool
Ble	Bleomycin resistance protein.
BMP	Bone morphogenic protein
BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
cDNA	Complementary DNA
CDRs	Complementarity determining region
CFP	Cyan Fluorescent Protein
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate.
CHD	Cytokine receptor homology domain
CHO	Chinese hamster ovary
CID	Collision-induced dissociation
CMT	*C*-mannosyltransferase
CRISPR	Clustered regularly interspaced short palindromic repeats
DDM	*n*-Dodecyl β-D-maltoside and
DK	Dolichol kinase
DME	Dulbecco's modified Eagle's medium
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
Dol-P-Man	Dolichol phosphate mannose
Dol-P	Dolichol phosphate
Dopa	3,4-dihydroxyphenyl-L-alanine
Dox	Doxycycline
DPM	Dolichol phosphate mannose synthase
dsDNA	Double stranded DNA
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EC_{50}	Half maximal effective concentration
ECL	Enhanced chemiluminescence

ECM	Extracellular matrix
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbetamide
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted ion chromatogram
ELISA	Enzyme-linked Immunosorbent assay
EPO	Erythropoietin
EPOR	Erythropoietin receptor
ER	Endoplasmic Reticulum
ERK	Extracellular signal regulated kinases
EtBr	Ethidium bromide
F5	Coagulation factor V
FACS	Fluorescence activated cell sorting
FCS	Fetal calf serum
FNIII	Fibronectin type III domain
Fru-6-P	Fructose-6- phosphate
FSC/SSC	Forward scatter/side scatter
FWR	Framework regions
G-CSF	Granulocyte colony-stimulating factor
GAGs	Glycosaminoglycans
GAP	Glyceraldehyde-3-phosphate dehydrogenase
gDNA	Genomic DNA
GDP-Man	GDP-D-mannose
GDP	Guanosine diphosphate
GFP	Green Fluorescent Protein
GHR	Growth hormone receptor
Glc-6-P	Glucose-6-phosphate
Glc	D-glucose
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GMPPA/B	GDP-mannose pyrophosphorylase (A/B)
Gp130	Glycoprotein 130 receptor
GPCR	G-protein-coupled receptors
GPI anchor	Glycosylphosphatidylinositol anchor
GTP	Guanosine triphosphate
HA	Hyaluronic acid
HCD	Higher-energy collision dissociation
HEK293	Human embryonic kidney 293 cell line
HEPEs	4-(2-hydroxyethyl)-1- piperazineethanesulfonic acid
HK	Hexokinase
HPLC	High-performance liquid chromatography
HRP	Horse radish peroxidase
Ig	Immunoglobulin
IL-12	Interleukin-12
IL-21R	Interleukin-21 receptor
IL	Interleukin
IL6R	Interleukin 6 receptor
IL7R	Interleukin-7 receptor
IMAC	Immobilised metal ion affinity chromatography
Indel	Insertion or deletion

IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyran
JAK	Janus kinase
kb	Kilobase
K_D	Dissociation constants
KLH	Keyhole limpet hemocyanin
KO	Knock-out
k_{off}	Dissociation rate
k_{on}	Association rate
KPS	KDN-9-phosphate synthase
L1CAM	L1 cell adhesion molecule
LB	Luria broth
LC-MS/MS	Liquid chromatography– tandem mass spectrometry
LCR	Low copy repeats
LDL	Low-density lipoprotein
LDLR	Low-density lipoprotein receptor
LEFTY1/2	Left-right determination factors 1 and 2
LPL	Lipoprotein lipase
LPS	Lipopolysaccharide
M-CSF	Macrophage colony-stimulating factor
mAb	Monoclonal antibody
MAC	Membrane attack complex
MAG	Myelin-associated glycoprotein
Man-6-P	Mannose-6-phosphate
Man	D-mannose
MAPK	Mitogen-activated protein kinases
MBP	Maltose-binding protein
MCS	Multiple cloning site
MEF	Mouse Embryonic Fibroblasts
MIC2	Microneme protein 2
MPI	Phosphomannose isomerase
MS	Mass Spectrometry
n	Number of experimental replicates
NFκB	Nuclear factor κ-light-chain- enhancer of activated B cells
NK	Natural killer cell
NMR	Nuclear magnetic resonance spectroscopy
OD	Optical Density
OGT	Octyl glucoside
OST	Oligosaccharyltransferase
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Poly-ethylene glycol
PGI	Phosphoglucose isomerase
PI	Propidium iodide
PIGZ	GPI mannosyltransferase 4
PM	Protein modifications

PMA	Phorbol 12-myristate 13-acetate
PMM2	Phosphomannomutase
POGLUT1	*O*-glucosyltransferase 1
PRLR	Prolactin receptor
q-PCR	Quantitative-polymerase chain reaction
RA	Rheumatoid arthritis
RAAS	Renin–angiotensin–aldosterone system
RMSD	Root-mean-square deviation
RNA	Ribonucelic acid
rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute
Rspo1	R-spondin-1
RT	Room Temperature
S2	*Drosophila* Schneider 2 cells
SB	Luria Super Broth
SD	Standard deviation
SDS	Sodium Dodecyl Sulfate
SEC	Size exclusion chromatography
SEM	Standard error of the mean
sf21	*Spodoptera frugiperda* cells
sGP	Ebola Virus soluble glycoprotein
sgRNA	Single guide RNA
SPR	Surface plasmon resonance
STAT	Signal transducer and activator of transcription
T-ALL	T-progenitor acute lymphoblastic leukaemia
THP1	Human monocytic AML cell line
TLR	Toll-like receptor
T_m	Melting temperature
TMHs	Transmembrane helices
TNF	Tumour necrosis factor
TPO	Thrombopoietin
TPOR	Thrombopoietin receptor
TRAP	Thrombospondin-related adhesive protein
TSR	Thrombospondin type-1 repeat domains
UNC5	UNC5 netrin receptor
UNC5A	UNC5A netrin receptor
UNC5B	UNC5B netrin receptor
UNC5C	UNC5C netrin receptor
UV	Ultra-violet
V	Volts
VEGFA	Vascular endothelial growth factor A
VEGFD	Vascular endothelial growth factor D
VWF	Willebrand factor
WB	Western-Blot
WT	Wild-Type
ϒc receptor	Common

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 1

Chapter 2

Chapter 5

Chapter 8

Table of Contents

Chapter 1

An introduction to tryptophan *C*-mannosylation

1.1 - Protein glycosylation

Protein modifications (PMs), which can occur during or after translation, enable the cell to expand and diversify its proteome, augmenting protein function and regulation beyond what is directly encoded by the genome. PMs involve the covalent chemical modification of proteins and are accompanied with changes in protein physical and chemical properties that influence its folding, trafficking, localisation, turnover, functional activity, and conformational dynamics [1]. Some of the most common PMs are glycosylation, phosphorylation, ubiquitination, acetylation and methylation [2]. Among the different types of PMs, protein glycosylation is the most diverse and abundant [3] [4]. It is estimated that around 65% of all proteins in humans undergo some form of protein glycosylation and more than 90% of cell surface proteins are glycosylated [3] [1]. Protein glycosylation can occur co- and post-translationally and involves the attachment of carbohydrate molecules to a protein.

In eukaryotes, all glycosylation occurs in either the endoplasmic reticulum (ER) or Golgi compartments, with the exception of the addition of *O*-GlcNAc moieties, which occurs in the cytoplasm [5]. As such, most glycoproteins transit through the secretory pathway and are either secreted, displayed on the cell surface or reside within the organelles of the cell [5]. The most common types of protein glycosylation in eukaryotes are *N*-glycosylation and *O*-glycosylation, with *C*-glycosylation being somewhat rarer [4]. *N*-glycosylation involves the attachment of a carbohydrate molecule to the side-chain nitrogen of an asparagine or, in the case of some bacteria, arginine [6]. *O*-glycosylation, involves the attachment of a sugar moiety onto the oxygen of a hydroxyl group of a serine, threonine, tyrosine, hydroxylysine or hydroxyproline residue [7]. The last and most unusual type of glycosylation is *C*-glycosylation, whereby a mannose is attached to the side-chain of a tryptophan residue forming a carbon-carbon bond [8].

1.2 - Tryptophan *C*-mannosylation

Tryptophan *C*-mannosylation involves the attachment of an α-D-mannopyranosyl unit onto the C2 carbon of the indole ring of tryptophan to form an unusual *C*-glycosidic bond (Figure 1.1A), which, in chemical terms, is radically different and far more stable than an *N*- or *O*-glycosidic bond [9] [8]. *C*-mannosylation was first discovered on RNaseII in humans in 1994, and there are now over twenty human proteins for which there is direct evidence of this modification [8]. Tryptophan *C*-mannosylation is predominantly found on proteins possessing a thrombospondin type-1 repeat (TSR) domain and on the ectodomains of the type-I cytokine receptors [10-12].

The glycosylation consensus sequence (sequon) for *C*-mannosylation is WXXW, with the modification occurring predominantly on the first tryptophan in the consensus sequence [13]. This sequon was established soon after the modification's discovery on RNaseII. This consensus sequence for *C*-mannosylation was established using alanine scanning across the first 13 residues of mature RNaseII. This experiment revealed that the only essential residues for *C*-mannosylation were W7 and W10, thereby establishing the WXXW motif of RNaseII as the minimum requirement for *C*-mannosylation. Identifying the WXXW sequon for *C*-mannosylation enabled the identification of over 350 human proteins as candidates for *C*-mannosylation [13].

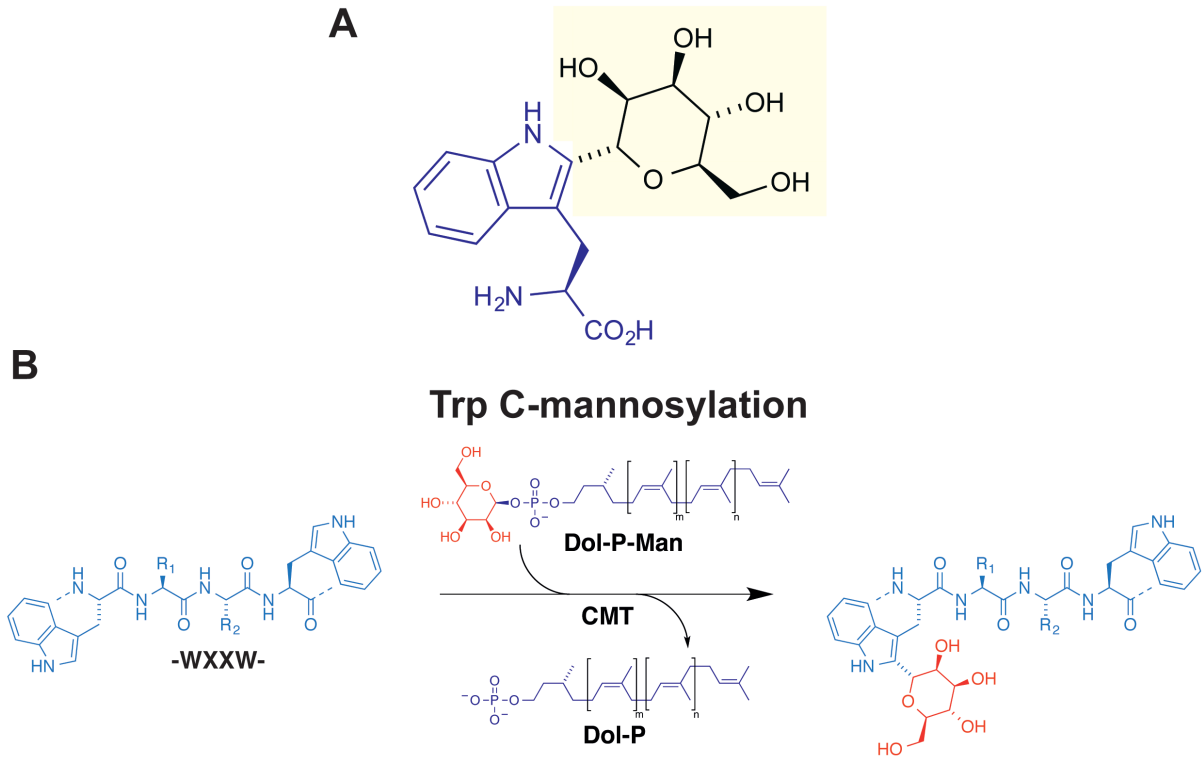


Figure 1. 1: The biochemistry of tryptophan *C*-mannosylation

a) Chemical structure of C^2-α-D-mannosyl-tryptophan

b) Biochemical reaction that takes place in the ER when a *C*-mannosyltransferase (CMT) transfers a mannose from a Dol-P-Man onto a tryptophan molecule of a target protein.

In 1998, several years after *C*-mannosylation was discovered on RNaseII, Doucey and colleagues discovered that *C*-mannosylation occurs in the ER, that it was catalysed by a specific unidentified *C*-mannosyltransferase (CMT), and that this enzyme utilised dolichol phosphate mannose (Dol-P-Man) as the sugar donor substrate (Figure 1.1B) [14]. Dol-P-Man is used in the ER as a sugar donor substrate for *O*-glycosylation, to assemble the lipid-bound oligosaccharide utilised by oligosaccharyltransferases (OST) in *N*-glycosylation, and in GPI anchor biosynthesis, making it a highly important metabolite for cellular glycosylation pathways [15]. Dol-P-Man is, in turn, biosynthesised on the cytoplasmic side of the ER membrane from dolichol phosphate (Dol-P) and GDP-D-mannose (GDP-Man), which is tied to the tightly regulated metabolism of mannose within the cell. Figure 1.2 illustrates the biosynthetic pathway for Dol-P-Man and mannose metabolism in the cell [15].

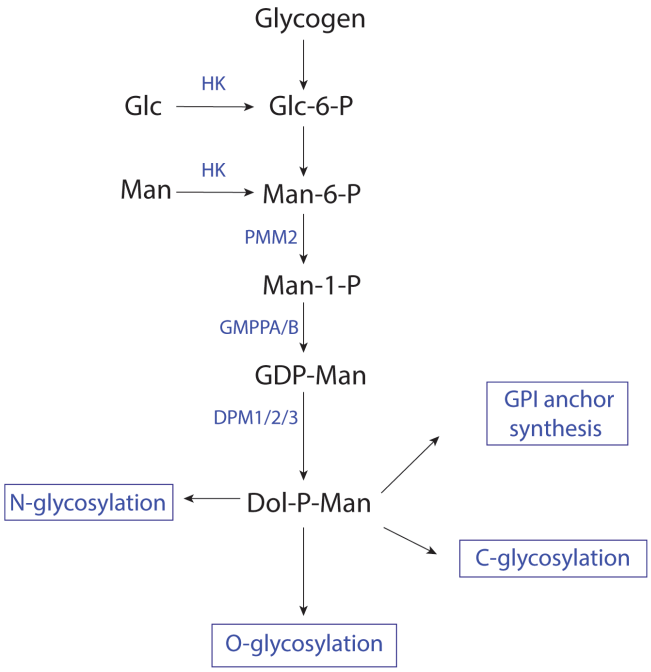


Figure 1. 2: Mannose metabolism and utilisation in eukaryotic cells.

Man, D-mannose; Glc, D-glucose; HK, hexokinase; MPI, phosphomannose isomerase; PMM2, phosphomannomutase; GMPPA/B, GDP-mannose pyrophosphorylase (A/B); PGI, phosphoglucose isomerase; KPS, KDN-9-phosphate synthase; Man-6-P, mannose-6-

phosphate; Fru-6-P, fructose-6- phosphate, Glc-6-P, glucose-6-phosphate; GDP-Man, GDP-mannose; Dol-P-Man, dolichol phosphate mannose; DPM1/2/3, Dol-P-Man synthase.

Most of what is known about the biosynthetic pathway of *C*-mannosylation is extrapolated from other well-studied glycosylation pathways such as *O*-mannosylation, *N*-glycosylation and GPI-anchor synthesis because they all share Dol-P-Man as the sugar donor substrate [16]. The biosynthetic pathway for *C*-mannosylation begins on the cytosolic surface of the ER membrane where a phosphate molecule from ATP is transferred to the membrane bound dolichol molecule by dolichol kinase (Figure 1.3) [16]. A mannose molecule is then transferred from GDP-Man onto the dolichol-phosphate molecule by dolichol phosphate mannose synthase to give Dol-P-Man. The Dol-P-Man molecule is then flipped from the cytosolic surface of the ER membrane into the ER lumen by a flippase [16]. From there, a *C*-mannosyltransferase transfers the mannose moiety from the Dol-P-Man onto the tryptophan of the WXXW motif of a nascent, presumably unfolded or misfolded, polypeptide substrate.

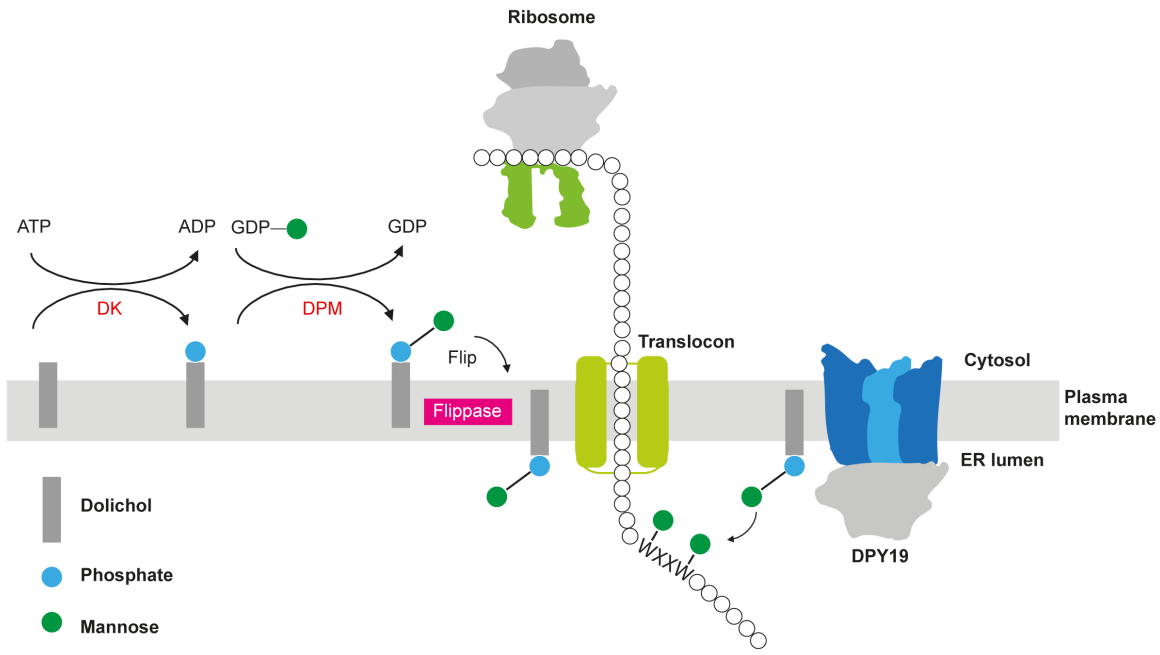


Figure 1. 3: Biosynthetic pathway of *C*-mannosylation in mammalian cells.

The pathway is initiated at the cytosolic surface of the ER membrane, and is terminated in the lumen of the ER. ADP, Adenosine diphosphate; ATP, Adenosine triphosphate; GDP, guanosine diphosphate; DK, dolichol kinase; DPM, dolichol phosphate mannose synthase.

1.3 - The tryptophan ladder and putative roles for tryptophan *C*-mannosylation

As mentioned above, proteins with TSRs and type-I cytokine receptors are commonly *C*-mannosylated. The WXXW motif of both of these proteins is involved in a unique structural element known as a tryptophan ladder, which is characterised by π-cation interactions between two or more tryptophan residues and one or more intercalating arginine residues [17] [18]. The planar stacking arrangement of the guanidinium moiety of arginine side chains over the centre of the aromatic rings of tryptophan stabilises protein tertiary structure [19].

The similarity of these structure can be best appreciated when the Fibronectin Type III (FNIII) module of the extracellular domain of the prolactin receptor (PRLR) is superimposed on the TSR domain of thrombospondin-1 (Figure 1.4) [17]. The root-mean-square deviation (RMSD) of the carbon-backbone in these overlaid structures is only 0.39 Å [17]. These strikingly similar features from very dissimilar proteins serve as an elegant example of convergent evolution at a protein structure level. A recently-developed 19-amino acid miniature protein consisting only of a tryptophan ladder revealed the structural and biophysical dynamics of the structural motif [20]. Structural characterization of the miniature protein revealed a novel topology composed of a β-strand:loop:PPII-helix [20]. Through mutagenesis studies, the authors concluded that the structure is stabilized by π-cation interactions and that each π-cation interaction can stabilise the miniature protein by up to -5.5 kcal/mol [20]. This value is on par with the Gibbs free energy of hydrogen bond formation and even protein folding, suggesting that formation of the tryptophan ladders may be important for stabilisation of protein fold.

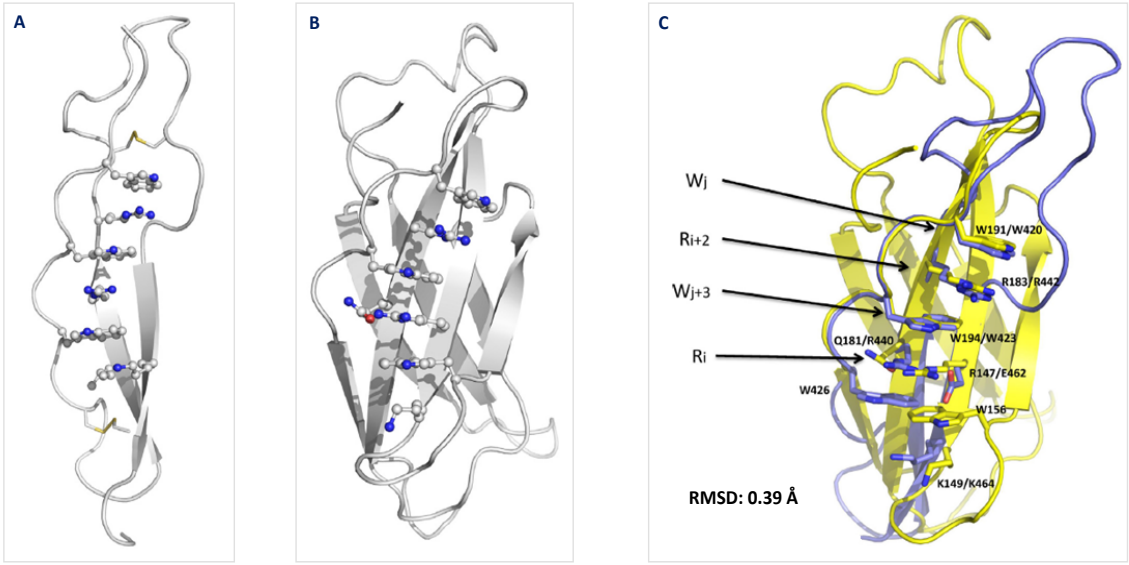


Figure 1. 4: Superposition of tryptophan ladders from cytokine receptors and TSRs.

a) TSR domain of Thrombospondin-1.
b) FNIII domain of the type I cytokine receptor, prolactin receptor (PRLR).
c) Superposition of the TSR and the type I cytokine receptor FNIII domains. (Adapted from Olsen and Kragelund) [17].

The roles of tryptophan *C*-mannosylation on these, or any other, protein remains to be definitely established. It is conceivable that *C*-mannosylation of the tryptophan residues in the WXXW motif help to assemble or stabilize the tryptophan ladder by pre-arranging the aromatic ring of tryptophan to encourage a π-cation interaction with the arginine side chain and shielding this hydrophobic amino acid from the bulk solvent [17]. Mutagenesis studies on Arg^{202} and Trp^{215} of the tryptophan ladder in the ϒc receptor (a type-I cytokine receptor) resulted in accumulation of the receptor in the ER: its failure to be exported to the plasma membrane suggests it was misfolded [21]. Another type-I cytokine receptor, IL21R, had the same phenotype when the key W and R residues in the tryptophan ladder were mutated [21]. While these studies speak to the importance of the tryptophan ladder, they cannot be used to infer a role for *C*-mannosylation. To examine the effects of *C*-mannosylation on the folding and stability of proteins with a tryptophan ladder, one must either produce the wild-type candidate proteins with and without *C*-mannosylation or examine their trafficking in cells with and without the tryptophan *C*-mannosyltransferase.

Another hypothesis for the role of *C*-mannosylation is that it plays a role in the function of proteins with tryptophan ladders, many of which bind sulphated glycosaminoglycans (GAGs), sulphatide, lipopolysaccharide (LPS) and phosphatidylserine. For example, the TSR domains of thrombospondin-1 are known to bind to heparin and heparin sulphate [22] and the tryptophan ladder formed by its KRFKQDGG**W**SH**W**SP**W**SS peptide mediates this interaction. However, when this TSR domain is expressed in *E. coli*, which does not produce *C*-mannosylated proteins [22], the heparin and heparin sulfate binding is ameliorated. The idea that *C*-mannosylation play a role in protein function is not mutually exclusive with the proposal that it also aids in protein folding and stabilisation.

1.4 - All of the human proteins with a predicted *C*-mannosylation sequon

Although the majority of the proteins discovered to date with a C^2-α-D-mannosyl-tryptophan have fallen in the type I cytokine receptor and thrombospondin type 1 repeat (TSR) domain families, there are many more protein families with the WXXW motif that are likely to bear the modification. Julenius constructed NetCGlyc, an online tool that predicts *C*-mannosylation sites in proteins based on the presence of a signal peptide and WXXW motif in their sequence [23]. A ScanProsite search using the WXXW motif indicates that there are >2000 proteins in the human proteome that contain this consensus sequence for *C*-mannosylation [24]. Further analysis of these proteins with TargetP and SignalP, which predict the sub-cellular location of these proteins, predicts if these sequences pass through the secretory pathway [25]. Approximately 500 of these proteins pass through the secretory pathway and are thus potentially *C*-mannosylated.

Gene ontology software can be used to predict what biological pathways these putatively *C*-mannosylated proteins are involved in [26] [27]. The heat map in Figure 1.5A shows the prevalence of these putatively modified proteins in different biological pathways. The most enriched pathway is the bone morphogenetic protein (BMP) signalling pathway, which senses a number of pleiotropic growth factors in different organs, not just bone [28]. BMPs have roles in osteogenesis and skeletogenesis, as well as embryogenesis and organogenesis [28]. There is also evidence that BMP signalling is involved in cellular differentiation and apoptosis [28]. Putatively *C*-mannosylated proteins are also highly enriched and may play a pivotal role in the immunity, particularly in: T-cell proliferation, the adaptive immune response and leukocyte cell-cell adhesion. This suggests that *C*-mannosylation plays a more important role in human biology than presently appreciated

These putatively *C*-mannosylated proteins belong to many different protein classes: not just TSR domains and type-I cytokine receptors (Figure 1.5B). The majority are cell surface receptors or enzymes, though some transporters, signalling molecules and secreted proteins also made the list. Of the receptors, only 19% were type-I cytokine receptors: the remaining 81% belonged to diverse receptor families that have not yet been definitely linked to *C*-mannosylation (Figure 1.6). G-protein coupled receptors account for 16% of these receptors and many of these receptors belong to the adhesion G-protein couple receptor sub-family (Figure 1.6). G-protein coupled receptors have a wide variety of molecular functions and transduced there downstream signalling either through cAMP or the phosphatidylinositol signal pathway [29]. An example is Adhesion G protein-coupled receptor B1(ADGRB1) which

is a cellular receptor that recognises phosphatidylserine and enhances the engulfment of apoptotic cells [30].

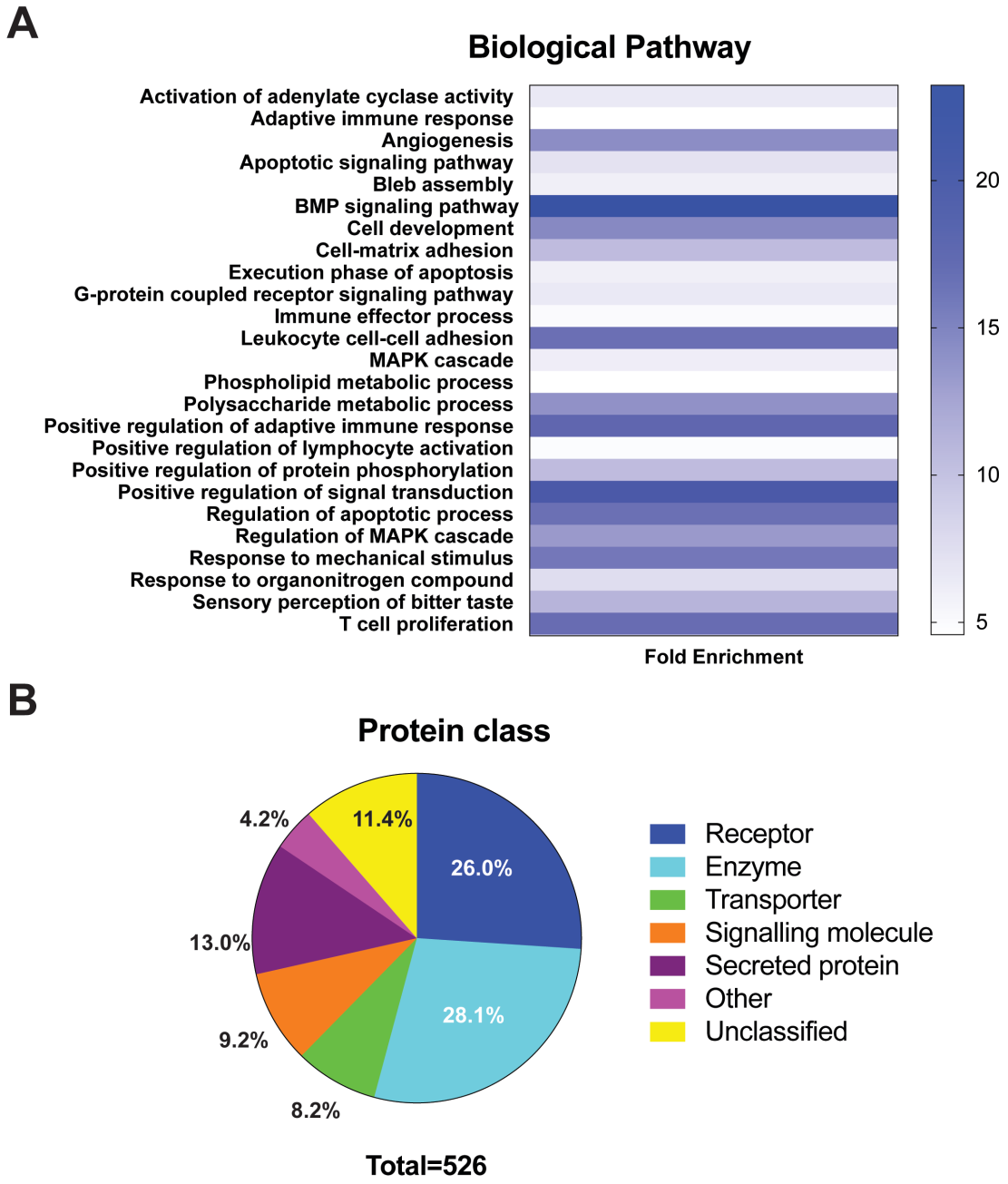


Figure 1. 5: Pathways and protein classes for putatively *C*-mannosylated human proteins.

a) Biological pathways replete with putatively *C*-mannosylated proteins.

b) Classification of putatively *C*-mannosylated proteins.

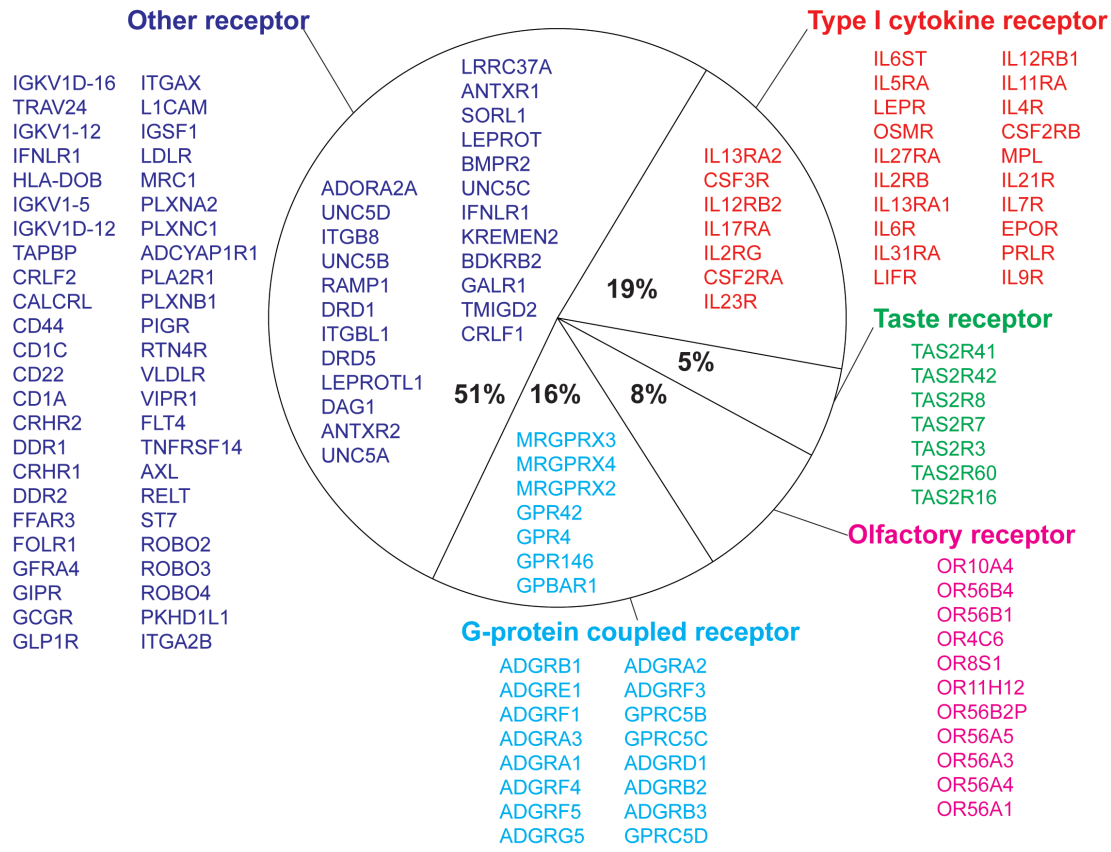


Figure 1. 6: Sub-classification of receptors with putative *C*-mannosylation sites.

Olfactory receptors also made the list. These receptors bind to odorant molecules in the nose to trigger a neuronal response which gives rise to the perception of smell (Figure 1.6) [31]. Taste receptors comprise 5% of the putative *C*-mannosylated receptor class, and one example is the TAS2R16 receptor which mediates the bitter taste response [32]. Approximately half of the receptors are a mixture of different family of proteins and examples include: LDLR, which is involved in the engulfment of cholesterol-rich LDL [33]; L1CAM, which is a receptor involved in cell adhesion and migration [34]; and the netrin receptors UNC5A, UNC5B, UNC5C and UNC5D, which contain a TSR domain and are involved in axon guidance, cell repulsion/attraction and apoptosis [35].

The majority of the enzymes with a putative *C*-mannosylation motif are hydrolases (57%) (Figure 1.7). Examples include: furin, an endoprotease that activates TGF-beta [36]; acetylcholinesterase (ACHE) and butyrylcholinesterase (BCHE), two enzymes that break down the neurotransmitters acetylcholine and butyrylcholine [37]; and angiotensin-converting enzyme 1 (ACE) and angiotensin-converting enzyme 2 (ACE2), which are involved in the

renin–angiotensin–aldosterone system (RAAS) and targets of commonly used ant-hypertension medications [38]. The other major class of enzymes that are putatively *C*-mannosylated are the transferases, which account for 30% of total enzymes (Figure 1.7). Many of these enzymes are glycosyltransferases such as: GPI mannosyltransferase 4 (PIGZ), which is an enzyme involved in glycosylphosphatidylinositol-anchor biosynthesis [39]; and protein *O*-glucosyltransferase 1 (POGLUT1), which transfers glucose and xylose from UDP-glucose and UDP-xylose, respectively, to a serine residue [40]. Oxidoreductases, kinases, ligases and isomerases are also represented in this list (Figure 1.7).

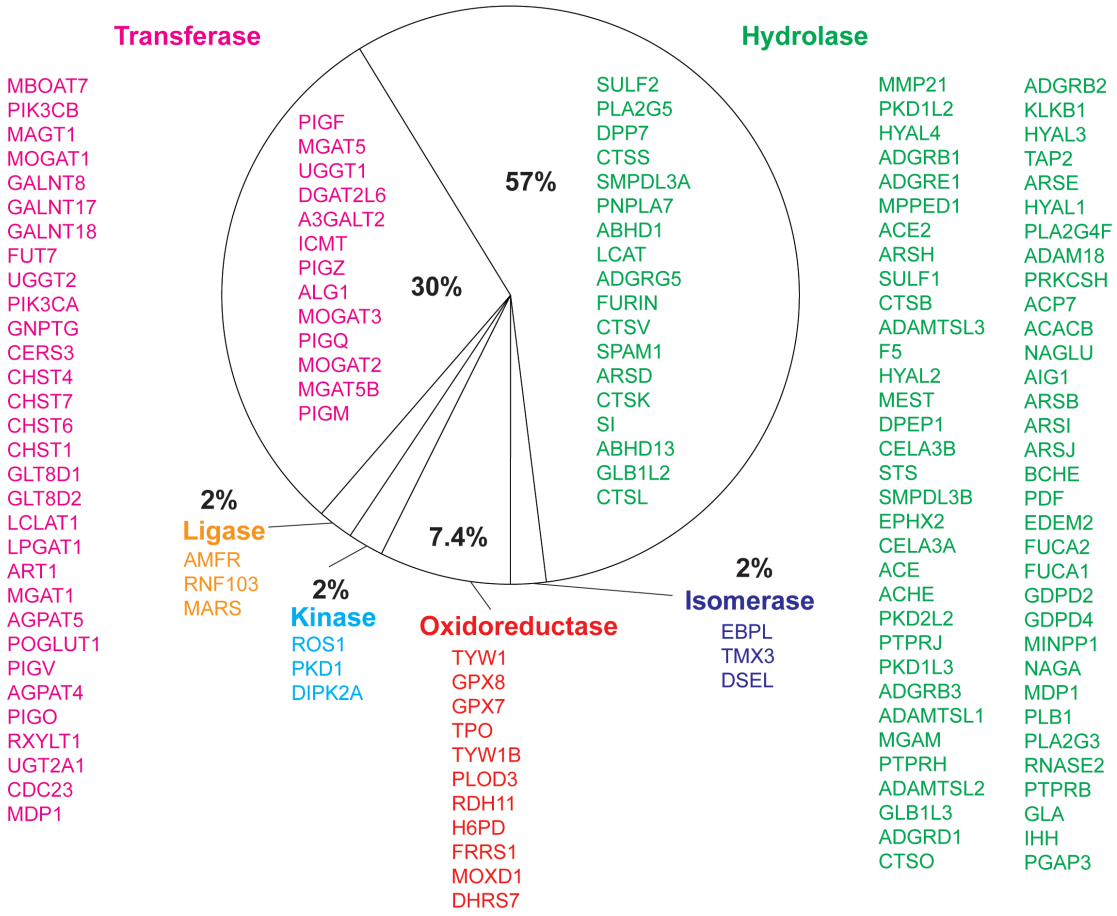


Figure 1. 7: Sub-classification of enzymes with putative *C*-mannosylation sites.

Of the signalling molecules with a putative *C*-mannosylation site, ≈60% are growth factors (Figure 1.8A). An overwhelming majority of these growth factors are bone morphogenetic proteins (BMP) (including all the GDF genes in the list). These BMPs a part of the transforming growth factor-beta (TGF-beta) protein family and are involved in a variety of different roles in the human body [28]. Other interesting growth factors include: vascular endothelial growth factor A (VEGFA) and vascular endothelial growth factor D (VEGFD) (Figure 1.8A), which

are glycoproteins involved in angiogenesis [41, 42]; and left-right determination factors 1 and 2 (LEFTY1/2), which are also members of the TGF-beta protein family and involved in left-right asymmetry determination [43]. The other signalling molecules are a mixed bag and include proteins like fibrinogen gamma chain (FGG), fibrinogen beta chain (FGB) and coagulation factor V (F5), which are all part of the blood coagulation system (Figure 1.8A). Other putative *C*-mannosylated proteins include enzyme modulators, calcium binding proteins and membrane traffic proteins (Figure 1.8B).

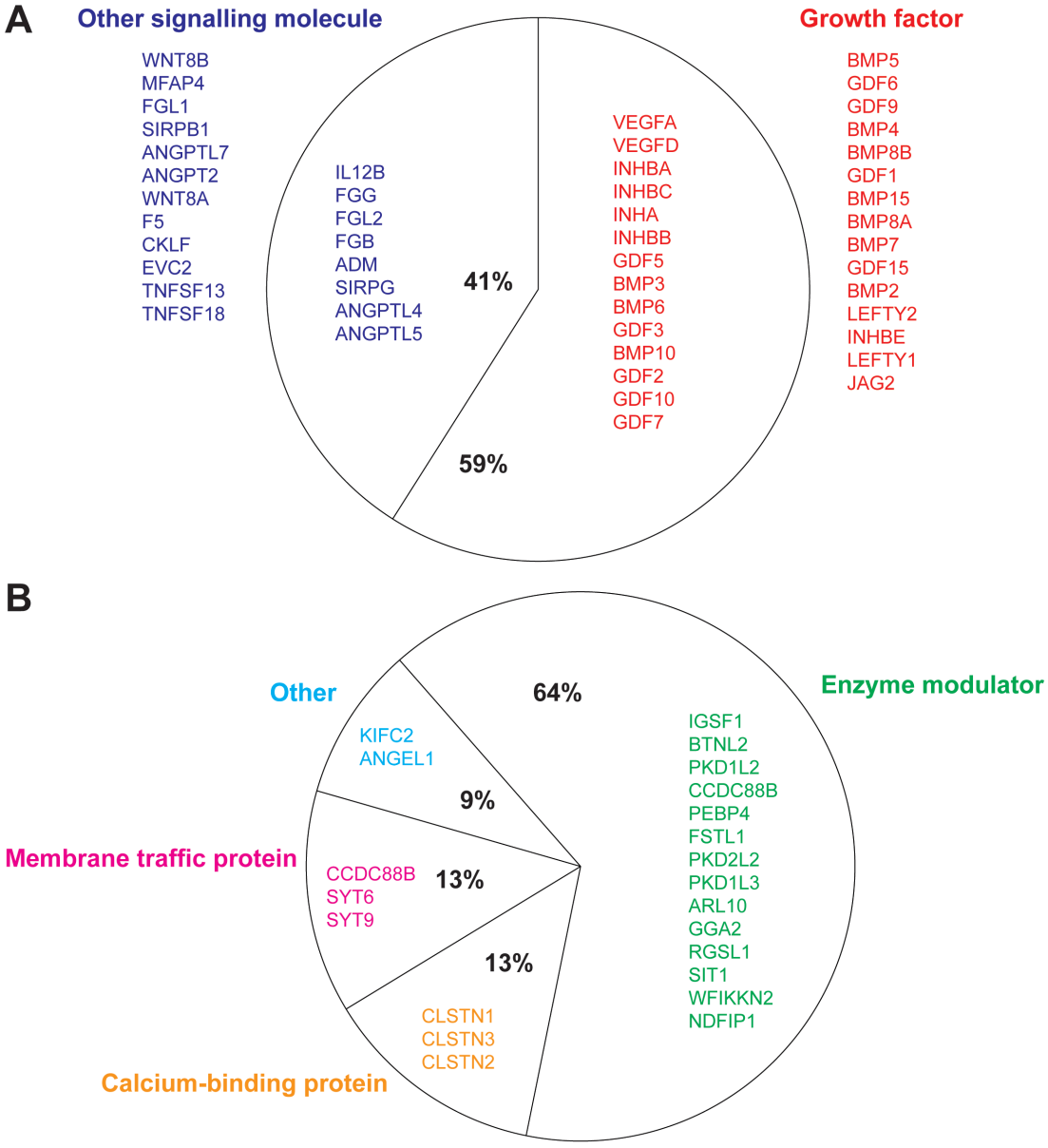


Figure 1. 8: Sub-classification of (a) signalling molecules and (b) other proteins with putative *C*-mannosylation sites.

Secreted proteins account for around 13% of all of the putatively *C*-mannosylated proteins identified in humans (Figure 1.9A). Within this group, ≈80% are classified as extracellular matrix proteins. Proteins in this class include the RSPON1-4 genes as well as SBSPON, which are all a part of the F-spondin family of proteins that regulate Wnt signalling in adult stem cells during development [44]. Another family of proteins is the ADAMTSL1-3 genes which are extracellular proteins involved cell-matrix interactions [45]. About 20% of the secreted proteins are cell adhesion molecules, such as transmembrane glycoprotein NMB (GPNMB), which plays a role in osteoblast differentiation and function (Figure 1.9A) [46].

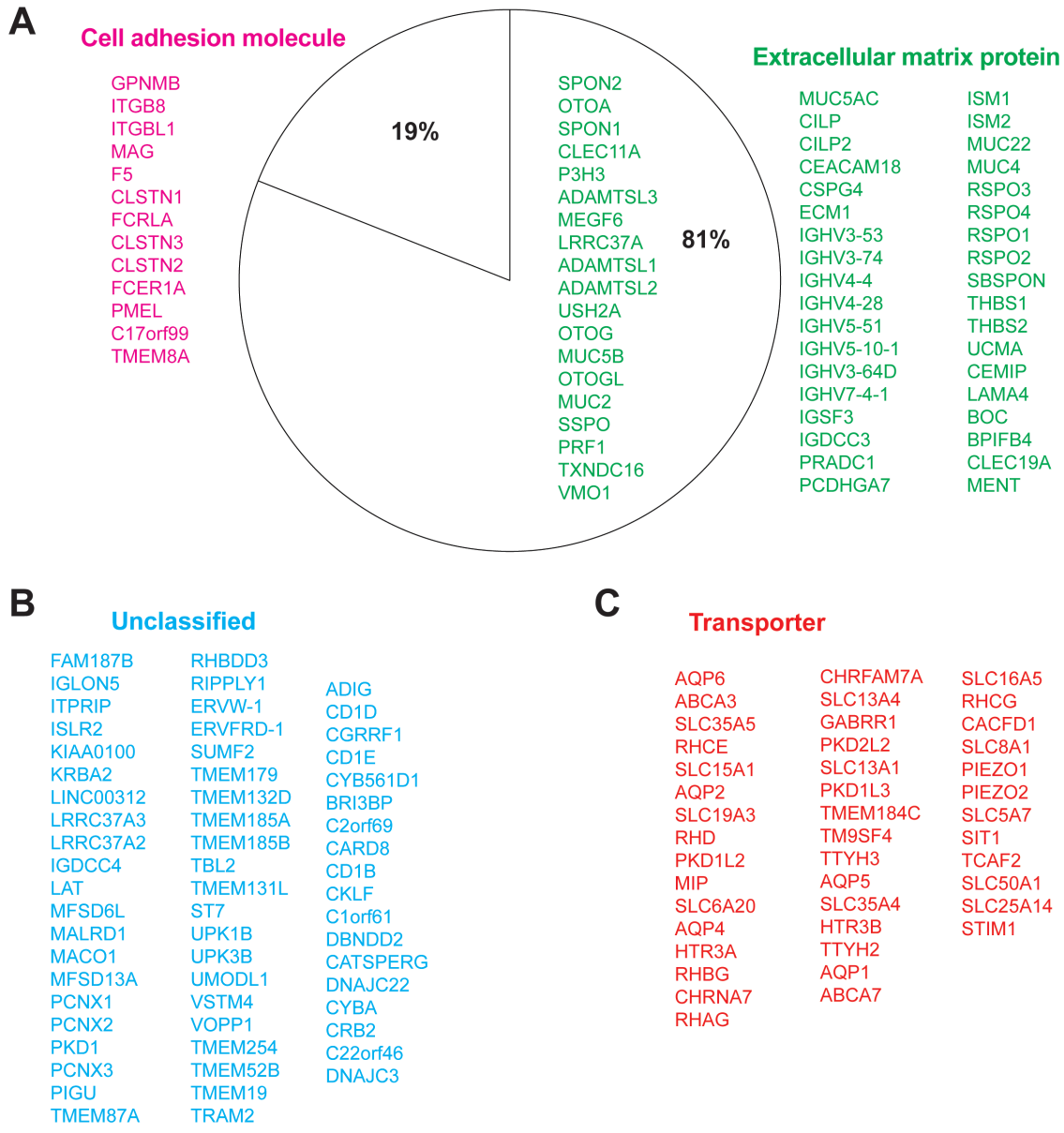


Figure 1. 9: Sub-classification of (a) secreted proteins, (b) transporters and (c) unclassified proteins with putative *C*-mannosylation sites.

The transporter class of proteins account for 8% of all of the putatively *C*-mannosylated proteins (Figure 1.9B) and includes the aquaporin gene family (AQP1-2, AQP4-6), which are integral membrane proteins that transport water in and out of cells [47]. The remaining 11% of the putatively *C*-mannosylated proteins did not have any assigned function (Figure 1.9B).

1.5 - Proteins for which there is direct empirical evidence of *C*-mannosylation

Table 1.1 lists a wide variety of proteins which have been proven to be *C*-mannosylated, mostly through mass spectrometry evidence. These proteins will be discussed in further detail in this chapter.

1.5.1 - Type-I cytokine receptors

Type-I cytokine receptors were initially identified and grouped together because they possessed a highly conserved WSXWS motif. This motif is consistent with the consensus sequence for *C*-mannosylation, WXXW, and so one might expect all of these proteins to be *C*-mannosylated. Type I cytokine receptors have similar structural features: they all contain an extracellular domain comprised of FNIII modules, a transmembrane domain and cytoplasmic signalling domain [48]. The extracellular domain always contains a cytokine receptor homology domain (CHD), which is composed of two FNIII modules [49]. One of the FNIII modules contains two pairs of conserved cysteine residues in the following configuration: CX-$_{(9–10)}$-CXWX-$_{(26–32)}$-CX-$_{(10–15)}$-C [49]. The second FNIII module contains the conserved WSXWS motif [50, 51]. The type I cytokine receptors can be characterised into five groups based on their structural homology, topology and evolutionary history (Figure 1.10) [48]. Group 1 receptor ectodomains consist solely of CHDs and form homodimers. They include the thrombopoietin receptor (TPOR, also known as c-mpl), prolactin receptor (PRLR), erythropoietin receptor (EPOR), CLF-3 receptor and the growth hormone receptor (GHR) [48]. Group 2 and group 3 receptors both contain a N-terminal Ig-domain on their extracellular domain but group 3 has a shortened intracellular domain [52] [53]. Group 2 and group 3 collectively comprise the IL6R family of receptor complexes. Group 4 receptors contain only an extracellular CHD domain and long intracellular domains [54] (Figure 1.10), whereas, group 5 receptors contain an Ig-like domain and short intracellular domains [54]. Group 4 and 5 chain receptors associate to form receptor complexes of the IL2R and IL3R receptor families (Figure

1.10). All of these possess the *C*-mannosylation WXXW sequon, with the exception of the growth hormone receptor (GHR).

Table 1. 1 List of *C*-mannosylated proteins

Protein Family	Species	Protein	Function
Thrombospondin type I repeats (TSRs)	*Homo sapiens*	ADAMTS-like 1/punctin-1	ECM component
	Homo sapiens	ADAMTS13	Blood coagulation
	Homo sapiens	Thrombospondin-1 (TSP-1)	ECM component
	Homo sapiens	Properdin	Complement system
	Homo sapiens	C6, C7, C8, and C9	Complement system
	Homo sapiens	Mindin	T-cell priming
	Homo sapiens	F-spondin	Neuronal development
	Homo sapiens	R-spondin-1	Wnt signalling
	Homo sapiens	RPE-spondin	Function unclear
	Homo sapiens	R-spondin-3	Wnt signalling
	Homo sapiens	UNC5A	Axon guidance
	C. elegans	MIG-21	Neuronal development
	C. elegans	UNC-5	Axon migration
Type I cytokine receptors	*Homo sapiens*	Interleukin-21 receptor	Immune system
	Homo sapiens	Thrombopoietin receptor	Blood development
	Homo sapiens	Erythropoietin receptor	Blood development
	Homo sapiens	Common gamma chain	Immune system
Other	*Homo sapiens*	Interleukin-12	Immune system
	Homo sapiens	RNaseII	RNase
	Homo sapiens	MUC5AC	Mucus component
	Homo sapiens	MUC5B	Mucus component
	Homo sapiens	Hyaluronidase 1	ECM component
	Homo sapiens	Lipoprotein lipase	Lipid metabolism
	Homo sapiens	Siglec-4	Nervous system
	Homo sapiens	EEF1A1	DNA translation
	C. morosus	Hypertrehalosaemic hormone	Neuropeptide
	Perna viridis	Pvfp-1 protein	Adhesive protein
	Bos taurus	MP20 lens protein	Cell organisation
	Ebola Virus	Soluble glycoprotein (sGP)	Unknown

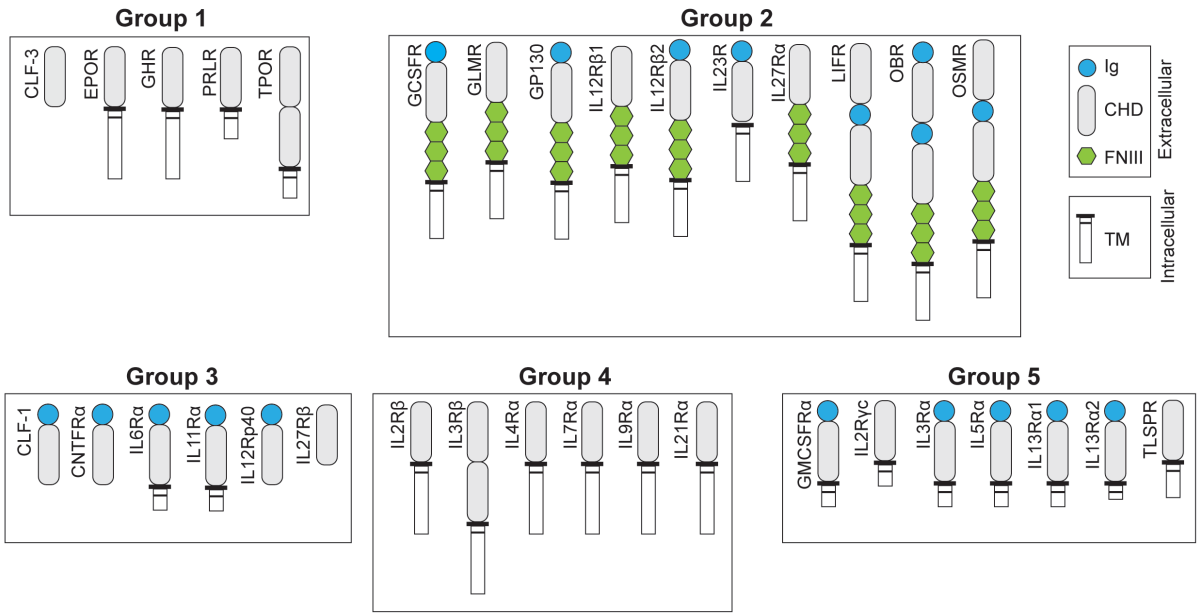


Figure 1. 10: Topological arrangement of type-I cytokine receptors.

Structural and topological arrangement of type-I cytokine receptors into five main groups.

1.5.1.1 - *C*-mannosylation of Interleukin 21 receptor

The Interleukin-21 receptor (IL-21R) is a type I cytokine receptor that is activated through the binding of Interleukin-21, which is a pleiotropic cytokine [55]. IL-21R is a heterodimeric receptor that is composed of an IL-21R α chain and a γc chain. Recently, the structure of the extracellular domain of the IL21-R bound to IL21 was determined and C^2-α-D-mannosyl-tryptophan identified on the protein's tryptophan ladder [12]. In this structure, the *C*-mannose appeared to play a role in multiple intra- and -inter molecular interactions through a hydrogen-bonding network between the arginine residues in the tryptophan ladder and a remote *N*-glycan from another FNIII domain of IL21R [12]. This supports the hypothesis that *C*-mannosylation may play a role in stabilising the π-cation network of the tryptophan ladder and the idea that mannosylated tryptophan can facilitate interactions between the tryptophan ladder and ligands.

A follow-up study by Siupka *et al.* has demonstrated that *C*-mannosylation is not only important for the structure of IL21R, but also important for the folding and trafficking of Il-21R to the plasma membrane [21]. HA-tagged WT-IL21R and R182A, N54A, and W195A mutants of IL21R were heterologously expressed in HEK293T and HeLa cells. Using immunofluorescence confocal microscopy, they found that the WT-IL21R was expressed on the cell surface while the W195A and R182A mutants accumulated in the ER with no cell surface expression [21].

1.5.1.2 - *C*-mannosylation of the thrombopoietin receptor

The thrombopoietin receptor (TPOR) is involved in the proliferation of multipotent hematopoietic bone marrow stem cells, their differentiation into mature megakaryocytes, and the production of platelets [56]. TPOR is a homodimeric receptor and signalling through the TPOR is initiated by the binding of its ligand, thrombopoietin (TPO), which then initiates the JAK/STAT and MAPK signalling cascade [56]. TPOR is interesting because it contains two WSXWS motifs. Recently, work by Sasazawa *et al.* revealed through mass spectrometry that TPOR is *C*-mannosylated at five tryptophan residues [57]. On the first WSXWS motif, Trp^{269} and Trp^{272} are *C*-mannosylated while on the second WSXWS motif Trp^{474} and Trp^{477} are *C*-mannosylated [57]. The fifth residue to be *C*-mannosylated is Trp^{416}, which is not part of any WSXWS motif [57]. These cryptic *C*-mannosylation sites may mean that there may be more *C*-mannosylated proteins in the human proteome than predicted in above (see 1.4).

Sasazawa *et al.* examined if *C*-mannosylation impacts the downstream signalling function of TPOR. They expressed recombinant flag-tagged TPOR in CHO cells together with the following mutants: W269F, W272F, W474F, W477F and W416F [57]. Using flow cytometry, the authors observed a decrease in expression of these mutant receptors on the surface of the cells, suggesting that *C*-mannosylation plays a role in protein folding and trafficking. These cell lines were stimulated with TPO and examined for the production of phosphorylated-STAT3, STAT3, phosphorylated-ERK, and ERK2, which are all downstream signalling proteins involved in TPO signal transduction [57]. There was no expression of phosphorylated-STAT3 and phosphorylated-ERK in the W269F, W272F, W474F, and W477F mutants as compared to wild-type TPOR [57]. This suggested that JAK/STAT and MAPK signalling was not initiated by these mutants. Notably, the W416F mutant, which is not located in any WXXW tryptophan ladder motif, did generate phosphorylated-STAT3 and phosphorylated-ERK, which suggests that it was functionally competent [57].

1.5.1.3 - *C*-mannosylation of the Erythropoietin receptor

The erythropoietin receptor (EPOR) was the first type I cytokine receptor proven to be *C*-mannosylated [10]. It is a homodimeric type I cytokine receptor responsible for erythropoiesis, which is the production of red blood cells, and also involved in the terminal differentiation of erythroid progenitors [58]. When its ligand, Erythropoietin (EPO), binds to EPOR it initiates a downstream JAK/STAT signalling cascade [58]. Like other type I cytokine receptors, it contains a highly conserved WSXWS motif. Hilton *et al*. carried out mutagenesis studies on

the WSXWS motif of EPOR [59]. All the residues in this motif were systematically mutated and examined for receptor export to the plasma membrane and ligand binding to receptor [59]. When the authors mutated Trp^{232} and Ser^{233}, there was a 300-fold decrease in receptor cell surface expression and EPOR was instead retained in the ER [59]. The Trp^{232} mutant also bound EPO 100-fold less than the wild type [59]. When overexpressed in HEK-EBNA, EPOR was found by mass spectrometry to be *C*-mannosylated on Trp^{232} with an occupancy of 50% [10]. Collectively, this alludes to possible roles for *C*-mannosylation in the trafficking of EPOR and in EPO binding.

1.5.1.4 - *C*-mannosylation of Interleukin-12

Interleukin-12 (IL-12) is a pro-inflammatory heterodimeric type-I cytokine that is composed of a β (p40) subunit and α (p35) subunit [60]. In actual fact, the p40 subunit is a homologue of the type-I cytokine receptors and is perhaps better described as a soluble type-I-cytokine receptor. It possesses the conserved WXXW motif within a FNIII domain that is characteristic of the type-I cytokine receptors. Doucey *et al*. expressed recombinant IL-12 β (p40) in Chinese hamster ovary (CHO) cells and showed by mass spectrometry that it was *C*-mannosylated on Trp^{319} [61]. IL-12 is expressed in minute amounts in IL-12 secreting cells, making it challenging to obtain enough material for mass spectrometry on endogenous samples. As yet, there is no information available on the possible function of *C*-mannosylation in IL-12 expression or activity.

1.5.2 - Thrombospondin type I repeats

The thrombospondin type I repeats (TSRs) are small domains approximately 60 amino acids in length with six conserved cysteine residues that hold the protein together through the formation of disulphide bonds [62]. The disulphide bonds can be found in two different configurations [62]. Group 1 TSR domains have a bonding configuration of C1–C5, C2–C6, C3–C4, while group 2 TSR domains have a configuration of C1–C4, C2–C5, C3–C6 [62]. The crystal structure of the TSR domains from thrombospondin-1 revealed that TSRs are composed of three antiparallel strands (Figure 1.11A). Strand A is irregular, but strand B and C form a β-sheet. Strand A of TSRs contain a conserved stretch of tryptophan residues arranged as WXXW(XXW), where X equals any amino acid, which is of course the consensus sequence for *C*-mannosylation. Most TSR domains also contain a YXRXZ motif (where, Y and Z are usually, R, K or Q) on strand B, with the lysine side chain interdigitating the aromatic

tryptophan side chains from strand A to form a tryptophan ladder with stabilising π-cation interactions [19] [18] (Figure 1.11B). These features are obvious at an amino acid sequence level for TSR domains that are known to be *C*-mannosylated (Figure 1.12, Table 1.2)

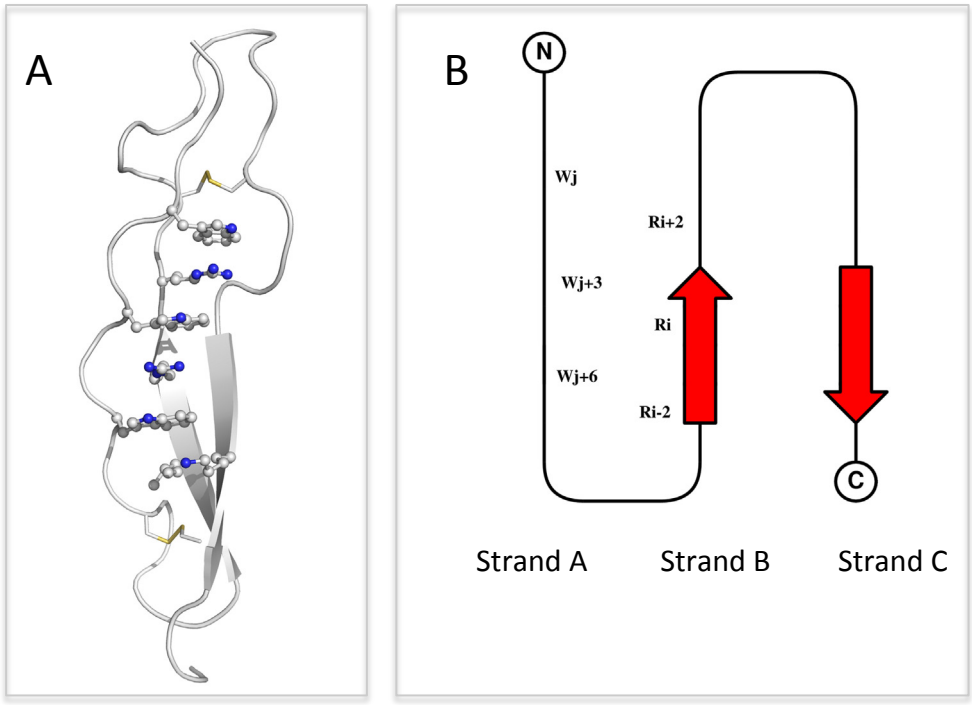


Figure 1. 11: (a) Structure and (b) configuration of TSR domains.

(Adapted from Olsen and Kragelund) [17].

```
TSR1-Properdin|     ---SPRWSLWSTWAPCSV-T------CSEGS----QLRYRRCVGWNGQ--CS-----GKVAPGTL-E--
TSR2-Properdin|     ---MGGWSGWGPWEPCSV-T------CSKGT----RTRRRACNHPAPK--CG-----G-HCPGQA----
TSR3-Properdin|     ---HGAWATWGPWTPCSA-S------CHGGPHEPKETRSRKCSAPEPSQKPP-----GKPCPGLA----
TSR4-Properdin|     ---AGGWGPWGPVSPCPV-T------CGLGQ----TMEQRTCNHPVPQ--HG-----GPFCAGDA----
TSR5-Properdin|     ---DGEWDSWGEWSPCIR-RNMKSISCQEIP--GQQSRGRTCRGRK----FD-----GHRCAGQQ----
TSR6-Properdin|     ---KGSWSEWSTWGLCMP-P------CGPNP---TRARQRLCTPLLPKYPPT-----VSMVEGQGEKNV
TSR1-C6|            --CFCDHYAWTQWTSCSK-T------CNSGT----QSRHRQIVVDKY---YQ-----ENFCEQICSK--
TSR2-C6|            ---NCLLGDFGPWSDCDP--------CIEK-----QSKVRSVLRPS---QFG-----GQPCTAPL----
TSR3-C6|            ---DGQWGCWSSWSTCDA-T---------YK----RSRTRECNNPAPQ--RG-----GKRCEGEK----
TSR1-C7|            ---NCQWDFYAPWSECNG--------CTKT-----QTRRRSVAVYG---QYG-----GQPCVGNA----
TSR2-C7|            ---DGGWSCWSSWSPCVQ-G---------K-----KTRSRECNNPPPS--GG-----GRSCVGET----
TSR1-C8alpha|       ---TCQLSNWSEWTDCFP--------CQDK-----KYRHRSLLQPN---KFG-----GTICSGDI----
TSR2-C8alpha|       ---DGSWSCWSSWSVCRA-G---------I-----QERRRECDNPAPQ--NG-----GASCPGRK----
TSR1-C8beta|        ---DCELSSWSSWTTCDP--------CQKK-----RYRYAYLLQPS---QFH-----GEPCNFSD----
TSR1-C8beta|        ---DGKWNCWSNWSSCSG-R---------R-----KTRQRQCNNPPPQ--NG-----GSPCSGPA----
TSR1-C9|            ---DCRMSPWSEWSQCDP--------CLRQ-----MFRSRSIEVFG---QFN-----GKRCTDAV----
TSR1-TSP-1|         ---DDGWSPWSEWTSCST-S------CGNGI----QQRGRSCDSLN-----------NRCEGSS----
TSR2-TSP-1|         ---DGGWSHWSPWSSCSV-T------CGDGV----ITRIRLCNSPSPQ--MN-----GKPCEGEA----
TSR3-TSP-1|         ---NGGWGPWSPWDICSV-T------CGGGV----QKRSRLCNNPTPQ--FG-----GKDCVGDV----
TSR1-ADAMTS13|      ---HGRWSSWGPRSPCSR-S------CGGGV----VTRRRQCNNPRPA--FG-----GRACVGAD----
TSR1-ADAMTS-L1|     ---DGLWDAWGPWSECSR-T------CGGGA----SYSLRRCLS-------------SKSCEGRN----
TSR1-Mindin|        ---DCEVSLWSSWGLCGGHC------GRLGT----KSRTRYVRVQP----AN-----NGSPCPEL----
TSR1-F-SPONDIN|     ---TCIYSNWSPWSACSSST------CDKGK----RMRQRMLKAQL----DL-----S-VPCPDT----
TSR2-F-SPONDIN|     ---TCTMSEWITWSPCSI-S------CGMGM----RSRERYVKQFP----ED-----GSVCTLPT----
TSR3-F-SPONDIN|     ---SCLMTEWGEWDECSA-T------CGMGM----KKRHRMIKMNP----AD-----GSMCKAET----
TSR4-F-SPONDIN|     ---PCLLSPWSEWSDCSV-T------CGKGM----RTRQRMLKSLA----EL-----G-DCNEDL----
TSR5-F-SPONDIN|     ---DCELTEWSQWSECNK-S------CGKGH----VIRTRMIQMEP----QF-----GGAPCPET----
```

Figure 1. 12: Sequence alignment of *C*-mannosylated TSR domains from human proteins.

The amino acid sequences of TSR domains from human proteins were aligned with Clustal Omega. Residues in yellow represent tryptophan residues in a WXXW motif. W in bold are known sites of *C*-mannosylation. Residues in red are conserved cysteine residues. Residues in grey represent the YXRXZ motif (where Y, Z = R, K or Q).

Table 1. 2: TSR domains with empirical evidence of *C*-mannosylation.

Protein	**Function**	**Organism**	**Type of evidence**	**Reference**
ADAMTS-like 1/punctin-1	ECM component	*Homo sapiens*	Mass spectrometry	[63]
Thrombospondin-1 (TSP-1)	Cell-to-cell and cell-to-matrix interactions	*Homo sapiens*	Mass spectrometry	[64]
Properdin	Positive regulator of complement system activation (alternative pathway)	*Homo sapiens*	Mass spectrometry and Edman degradation	[65]
Complement component C6, C7, C8, and C9	Complement system (innate immunity)	*Homo sapiens*	Mass spectrometry and NMR spectroscopy, plus structure of C6	[11, 66]
UNC-5	Axon migration	*C. elegans*	Mass spectrometry	[67, 68]
Mindin	Extracellular matrix protein required for efficient T-cell priming by dendritic cells.	*Homo sapiens*	Mass spectrometry	[69]
MIG-21	Left-right asymmetric polarization of the Q neuroblasts.	*C. elegans*	Mass spectrometry	[68]
F-spondin	Involved in neural cell pattern and axonal growth	*Homo sapiens*	Mass spectrometry	[70]
R-Spondin-1	Activator of Wnt signalling pathway	*Homo sapiens*	Mass spectrometry	[71]
RPE-spondin	Function unclear	*Homo sapiens*	Mass spectrometry	[72]
R-spondin-3	Activator of Wnt signalling pathway	*Homo sapiens*	Mass spectrometry	[73]
UNC5A	Axon guidance, cell repulsion/propulsion.	*Homo sapiens*	Mass spectrometry	[74]

1.5.2.1 - *C*-mannosylation of mindin

Mindin (Spondin-2) is an extracellular matrix protein which belongs to the F-spondin protein family and has a variety of different functions [75-80]. Its primary function is to serve as a pathogen recognition molecule where it recognises LPS on bacterial cell surfaces and serves as an opsin for phagocytosis by macrophages [81]. Mice lacking mindin cannot clear pathogens and macrophages deficient in mindin show defective responses to microbial stimuli [81]. Mindin is composed of a TSR domain and a spondin domain, the spondin domain is known to bind to $\alpha_M\beta_2$ integrin on macrophage cell surfaces [69]. Currently, there is no information as to which of the two domains mediate the binding to LPS, or if they are both required for binding. Mindin is *C*-mannosylated on both W^{283} and W^{286} within the TSR domain (aa277-aa331) [69]. Full-length mindin produced in mammalian cells showed significant binding to LPS, but full-length mindin produced in bacterial cells, or the spondin domain alone, showed only weak binding to LPS [69]. This information suggests that *C*-mannosylation of the mindin TSR domain is important for LPS binding.

1.5.2.2 - *C*-mannosylation of properdin

Properdin is a glycoprotein that is a positive regulator of the alternative pathway of the complement system [82]. The complement system is an important part of the innate immune system that acts as an immediate first line defence against pathogens or foreign bodies [83]. Properdin is a highly *C*-mannosylated protein comprised of six TSR domains. Out of the seventeen-tryptophan residues in the six WXXW(XXW) motifs, fourteen tryptophan residues are *C*-mannosylated [65]. The impact of *C*-mannosylation on properdin expression and function remain unknown.

1.5.2.3 - *C*-mannosylation of complement components

The complement system acts as the first line of defence against microbial pathogens [83]. The formation of the membrane attack complex (MAC) and its subsequent binding to cell membranes of pathogens causes cell lysis and elimination of the pathogens [84]. Complement components C5b C6, C7, C8, and C9 form a tandem repeat polymer that forms the MAC of the innate immune system [84]. Complement component C6 has three TSR domains, C7 has two TSR domains, while C8α and C8β contain two TSR domains each and C9 contains only one TSR domain. Mass spectrometry studies by Hofsteenge *et al*. found that C6, C7, C8α, C8β and C9 were all *C*-mannosylated on multiple tryptophan residues [11]. C6 is modified on Trp^8

and Trp^{11} located on TSR1, Trp^{69} which is not located on any TSR domain, and Trp^{547}, Trp^{550} and Trp^{553} located on TSR3 [11]. C7 is modified on Trp^{14} not located on any TSR domain and on Trp^{481}, Trp^{484} and Trp^{487} located on TSR2. C8α is modified on Trp^{14} not located on any TSR domain, and on Trp^{512}, Trp^{515} and Trp^{518} located on TSR2 [11]. C8β is modified on Trp16 and Trp^{19} on TSR1, and Trp^{497} and Trp^{500} located on TSR2 [11]. C9 is modified on Trp^{27} and Trp^{30} on TSR1 [11]. A crystal structure of C8 has been solved and reveals electron density for these C^2-α-D-mannosyl-tryptophans [66], which cluster together from both the C8a and C8b subunits.

1.5.2.4 - *C*-mannosylation of ADAMTS13

ADAMTS13 is a plasma metalloproteinase and a member of the ADAMTS family (a disintegrin-like and metalloproteinase with thrombospondin type 1 repeats) of proteins [85]. Its main function is regulating blood coagulation, where it cleaves Willebrand factor (VWF) multimers which attach platelets to endothelial cells thereby regulating platelet adhesion and aggregation [86]. It was shown by mass spectrometry that plasma-purified (endogenous) ADAMTS13 was *C*-mannosylated on Trp^{390} and Trp^{884} in TSR 1 and TSR4, respectively [87]. Another study, showed that the WXXW motif in TSR 1 of ADAMTS13 was important for its secretion and proteolytic activity by making alanine mutations of the first tryptophan residue [88]. This study, unlike other studies looking at mutants of the WXXW motif, did not make unfounded claims that *C*-mannosylation was responsible for the lack of secretion or activity in the WXXW mutant, though it does, once again, highlight the importance of the tryptophan ladder for protein folding and trafficking.

1.5.3 - Other proteins classes with evidence of *C*-mannosylation

1.5.3.1 - *C*-mannosylation of mucins

Mucins comprise a family of large, complex and heavily glycosylated glycoproteins that are characterised by their ability to form gels. Their functions range from lubrication, formation of chemical barriers and cellular signalling and are implicated in lung diseases such as asthma, bronchitis, chronic obstructive pulmonary disease (COPD) and cystic fibrosis [89] [90, 91]. MUC5AC and MUC5B are mucins that have been reported to be *C*-mannosylated based solely on lectin binding [92]. The cysteine-rich domains (CysDs) of MUC5AC and MUC5B possess

a WXXW motif and, when this domain was expressed in COS-7 cells, the protein could be captured by mannose-binding lectins (MBL). When the authors mutated the first tryptophan in the WXXW motif, the MBLs were not able to capture the domains, which was interpreted as the proteins no longer being *C*-mannosylated [92]. Immunofluorescence studies also suggested that mutation of the WXXW motif in mucin CysDs led to the protein being retained in the ER [92]. Expression of mucin CysDs in CHO-Lec35.1 cells, which have low Dol-P-Man levels, blocked secretion of the wild-type mucin CysD [92]. However, since Dol-P-Man is required for *N*-and *O*-glycosylation it is difficult to draw meaningful conclusions from this data. A recent study has also mutated the first tryptophan of the WXXW motif into an alanine in the MUC5B CysD and showed that this mutant could not be secreted, again, the authors extrapolated from this to erroneously conclude that *C*-mannosylation was essential for folding and trafficking of this mucin CysDs [93]. A continuing problem in the literature is the over interpretation of WXXW mutagenesis studies: removing a tryptophan that is important for a structural motif is not the same as preventing mannosylation of that motif [20]. On balance, claims that mucin CysDs are *C*-mannosylated remain contentious, as there has been no direct observation of a *C*-linked hexose.

1.5.3.2 - *C*-mannosylation of hyaluronidases

In humans there are five hyaluronidase homologues that break down hyaluronan (hyaluronic acid, HA), which is a non-sulphated glycosaminoglycan (GAG) comprised of alternating glucuronic acid and *N*-acetyl glucosamine [94]. Hyaluronan is found in the extracellular matrix, cartilage, epithelial, connective and neural tissues [94] and controls fluid homeostasis in these tissues [95]. Mass spectrometry experiments have demonstrated that Hyaluronidase 1 is *C*-mannosylated [96]. The WXXW motif modified in hyaluronidase 1 is conserved in the other four human hyaluronidases, indicating that all of these enzymes are probably *C*-mannosylated.

1.5.3.3 - *C*-mannosylation of lipoprotein lipase

Mass spectrometry has recently revealed that lipoprotein lipase (LPL) is *C*-mannosylated on W^{417} when recombinantly expressed in HT1080 cells [97]. LPL is an enzyme that is involved in lipid transport and metabolism where it is responsible for the hydrolysis of core triglycerides in lipoproteins [98]. Mutation of this tryptophan to a phenylalanine residue resulted in a decrease in protein secretion and enzymatic activity, relative to the wild-type enzyme [97].

Whether or not the W to F mutation, the loss of *C*-mannosylation, or both, are responsible for this decreased expression and activity remains to be determined.

1.5.3.4 - *C*-mannosylation of myelin-associated glycoprotein (Siglec-4)

Myelin-associated glycoprotein (MAG), also known as siglec-4, is a type-I transmembrane glycoprotein, lectin and signal transducer. Its main function is in glial-axonal interactions, but also has roles in inhibition of nerve regeneration [99] [100]. It recognises alpha 2,3-*N*-acetylneuraminic on the terminal galactose of a gangliotetraose core [101]. X-ray crystal structures of recombinant MAG expressed in mammalian cells reveal a C^2-α-D-mannosyl-tryptophan on W^{22} [102], which was confirmed using mass spectrometry. This tryptophan is part of a WXXW motif that is not conserved in other siglecs. This glycosylation site is near the neuraminic acid binding site of siglec-4 and mutations of this residue actually increase the affinity of for neuraminic acid ligands. Thus, it may be that tryptophan mannosylation plays a role in negatively regulating the activity of siglec-4 (MAG) [102].

1.5.3.5 - *C*-mannosylation of hypertrehalosaemic hormone

Though the focus here has been on mammalian biology, tryptophan mannosylation is also prevalent in other metazoans. *C*-mannosylation has been identified in the stick insect *Carausius morosus* hypertrehalosaemic hormones, which are small neuropeptides that have a role in the regulation of the trehalose concentrations in insects [103]. This was contentious at the time because *Drosophila melanogaster* Schneider 2 cells do not demonstrate any *C*-mannosyltransferase activity [71] [68]. The same is true for the Sf9 and Sf21 cells derived from *Spodoptera frugiperda*. It may be that these lab-adapted, immortalised cell lines simply do not express their *C*-mannosyltransferase. The hypertrehalosaemic hormone from *Carausius morosus* was shown to be *C*-mannosylated on Trp^8 using NMR [104]. This complemented earlier mass spectrometry results demonstrating a hexose on one of the tryptophan residues of a similar peptide from a related species [105]. NMR diffusion measurements showed that the mannose reduced the hydrodynamics radius of this peptide as compared to the unmodified peptide and that the conformation of the modified peptide changed relative to the unmodified form, which was in equilibrium with higher-order aggregates [104]. This was the first and remains the best direct evidence that tryptophan *C*-mannosylation can impact protein/peptide folding.

1.5.3.6 - *C*-mannosylation of Pvfp-1, a mussel adhesive protein

An interesting case of *C*-mannosylation has been observed by mass spectrometry in the mussel *Perna viridis* with the Pvfp-1 protein, which is a mica-adhesive protein found with numerous *C*-mannosylated tryptophan residues [106]. Mussels contain multiple adhesive proteins that allow them to adhere to the rock substrates on which they grow. These adhesive proteins are dominated by the presence of 3,4-dihydroxyphenyl-L-alanine (Dopa), which is largely responsible for this adhesive property, as seen in Pvfp-5, a homologue of Pvfp-1 [107] [108]. In contrast to Pvfp-5 and other adhesive proteins, Pvfp-1 only has one Dopa amino acid in all of the protein but instead has 42 repeats of two decapeptides, $ATPKPW_1TAW_2K$ and $APPPAW_1TAW_2K$ in tandem [106]. Roughly 80% of this protein is composed of these two decapeptides repeats, and all tryptophan residues are modified by *C*-mannosylation [106]. While W_1 was a simple C^2-α-D-mannosyl-tryptophan, W_2 was also hydroxylated at C7 too (C^2-α-D-mannosyl-7-hydroxy-tryptophan) [106]. It has been hypothesized that the C^2-α-D-mannosyl-tryptophan is likely important for the ability of this peptide to bind aluminosilicates (e.g. mica) [106]. In another study, work was conducted on what properties of the C^2-α-D-mannosyl-tryptophan in Pvfp-1 were causing this highly adhesive property [108]. They asserted that the π-cation interactions between the C^2-α-D-mannosyl-tryptophan and lysine side chains provide a low desolvation penalty and that this facilitates conformational changes in the peptide at the surface of aluminosilicates and possibly explains why Pvfp-1 has such strong adhesive properties [108]. As such, *C*-mannosylated peptides may have applications in the development of new biocompatible materials and adhesives.

1.5.4.7 - *C*-mannosylation of MP20 lens protein

Mass spectrometry has provided evidence for *C*-mannosylation on the Bovine Lens fibre membrane intrinsic protein (lim2), or MP20 lens protein [109]. Lim2 is a member of the tetraspanin superfamily and is the second most abundant membrane protein in the mammalian eye lens fibre cells, which accounts for 99% of cell types in the lens [110]. Lim2 was identified as a ligand for Galectin-3, which is involved in cell-cell adhesion and cell-extracellular matrix interactions [110]. It is hypothesized that lim2 may play a role in the organisation of lens fibre cells in the lens of the eye [110]. However, the intriguing aspect of this protein is that it is *C*-mannosylated on two non-canonical *C*-mannosylation sites: W^{42}RYC and W^{61}NAT [109]. The occupancy at W^{42} was higher than for W^{61} [109]. Humans have a homologue of lim2, which is 90% identical to the bovine lim2 and these tryptophans are conserved (WRYC and WANT), and so it is quite likely that these non-canonical sites are mannosylated in humans too. Again,

this suggests that there are probably many examples in the human proteome of unexpected *C*-mannosylation events.

1.5.4.8 - *C*-mannosylation of viral proteins

Ebola Virus soluble glycoprotein (sGP) was the first viral protein to be identified as *C*-mannosylated. sGP was expressed in HEK293T cells and was found by mass spectrometry to be *C*-mannosylated on Trp^{288} within a WXXW motif [111]. W238F mutants of this glycoprotein had no defects in trafficking or processing. The functions of sGP are still not well understood; therefore, a full analysis of the impacts of *C*-mannosylation on sGP have not be explored. sGP is not the only viral protein to have the WXXW motif. A RefSeq search returns over 2000 viral proteins with the WXXW motif [111]. Table 1.3 details viral proteins with a WXXW motif that are targeted to the secretory pathway and thus likely to be *C*-mannosylated. These proteins include examples from the Hepatitis B virus, Human immunodeficiency virus (HIV), Ebola Virus, Influenza A and C viruses, SARS coronavirus, Hepatitis C virus and Dengue virus type 1. Like the type I cytokine receptors and TSR domains, the WXXW motif of these proteins is accompanied with YXRXZ motif, suggests that these viral proteins may also possess a tryptophan ladder (Figure 1.13). Structural data for these regions of these glycoproteins is not presently available: it will be interesting to see if this prediction proves to be true.

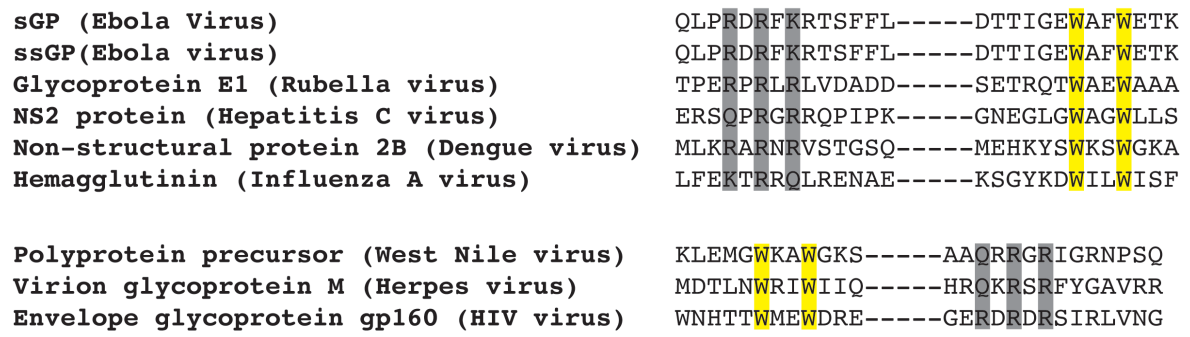


Figure 1. 13: Alignment of the putative tryptophan ladders of viral glycoproteins.

The amino acid sequences of the viral proteins were aligned using Clustal Omega. Residues in yellow indicate the tryptophan residues from the WXXW motif, and residues in grey indicate the YXRXZ motif (where Y, Z= R, K or Q).

Table 1. 3: Viral proteins with the WXXW *C*-mannosylation consensus motif.

Family	Virus	Protein
Poxviridae	Molluscum contagiosum virus	MC011L
Herpesviridae	Human herpesvirus 1	Virion glycoprotein M
	Human herpesvirus 2	Virion glycoprotein M
	Human herpesvirus 2	Tegument protein
Hepadnaviridae	Hepatitis B virus	S protein
Retroviridae	Human immunodeficiency virus	gp41
	Human immunodeficiency virus	gp160
Reoviridae	Human rotavirus	VP7
Bornaviridae	Borna disease virus	Glycoprotein
Rhabdoviridae	Vesicular stomatitis virus	Glycoprotein
	Mokola virus	Transmembrane glycoprotein
Filoviridae	Cote d'Ivoire ebolavirus	Secreted glycoprotein
	Cote d'Ivoire ebolavirus	Structural glycoprotein
	Reston ebolavirus	Secreted non-structural glycoprotein
	Reston ebolavirus	Structural glycoprotein
	Sudan ebolavirus	Secreted glycoprotein
	Sudan ebolavirus	Structural glycoprotein
	Zaire ebolavirus	sGP
	Zaire ebolavirus	ssGP
	Zaire ebolavirus	Virion spike glycoprotein precursor
Orthomyxoviridae	Influenza A virus	Hemagglutinin
	Influenza C virus	CM2 protein
Bunyaviridae	Hemorrhagic fever virus	Glycoprotein precursor
Coronaviridae	Human coronavirus 229E	Surface glycoprotein
	Human coronavirus 229E	4a protein
	Human coronavirus 229E	Membrane protein
	SARS coronavirus	E2 glycoprotein precursor
	SARS coronavirus	Matrix protein
Flaviviradae	Dengue virus type 1	Non-structural protein 2B
	West Nile virus	Polyprotein precursor
	Hepatitis C virus	NS2 protein
Togaviridae	Ross River virus	6K protein
	Rubella virus	Glycoprotein E1

1.6 - Discovery of the first *C*-mannosyltransferase

The genetic basis of *C*-mannosylation was unknown until 2013, when the Bakker lab reported that *dpy19* encoded a *C*-mannosyltransferase in *Caenorhabditis elegans* [68]. The rationale behind the discovery of the *C. elegans C*-mannosyltransferase was based on anticipated similarities between the enzymes that catalyse *N*-glycosylation and *C*-mannosylation [68].

Oligosaccharyltransferase (OST), the enzyme responsible for *N*-glycosylation, is a multi-pass transmembrane domain protein with seven transmembrane helices (TMHs) that resides in the ER and utilises a sugar dolichol phosphate donor substrate [112]. Since *C*-mannosylation requires Dol-P-Man, a sugar dolichol phosphate donor substrate, it was reasoned that it might share some sequence similarities with OST. Based on this assumption, *dpy19* in *C. elegans* was identified as a prime candidate. The name of this gene derives from DumPY-19 because mutations in this gene in *C. elegans* result in a shortened (dumpy) phenotype [113]. Heterologous expression of *C. elegans* DPY19 in Schneider 2 (S2) cells, which have no *C*-mannosyltransferase activity, and isolation of the ER-rich microsomal fractions enabled an *in vitro* demonstration of *dpy19*-dependent *C*-mannosyltransferase activity by way of a radio assay [68]. Co-expression of TSR domains from *C. elegans* UNC5 and MIG-21 in these cells, followed by mass spectrometry on the purified proteins, enabled direct detection, localisation and quantification of the *dpy19*-dependent *C*-mannosylation [68]. When MIG-21 was expressed in the parental S2 cells lacking *dpy19*, it was not secreted, suggesting that *C*-mannosylation is important for the folding and or secretion of MIG-21 [68].

1.7 - Conservation and evolution of the *dpy19* gene family.

The first mention of the *dpy19* gene in literature was in 2000 when it was discovered that the *C. elegans dpy19* gene was essential in QL and QR neuroblast polarisation [113]. The term *dpy19*, or DumPY-19, was given to these mutants because this gene caused a Q neuroblast migration defect that resulted in a dumpy phenotype (short and stout) [113]. The *dpy19* gene family encodes multi-pass integral membrane proteins and have undergone gene duplication and relocation events in both ancient and relatively recent times [114]. There are four *dpy19* homologues in mammals: *dpy19l1*, *dpy19l2*, *dpy19l3* and *dpy19l4*. Phylogenetic analyses by Carson *et al*. have traced the evolution of these genes across species (Figure 1.14) [114]. The first duplication events of the ancestral *dpy19* gene occurred before the divergence of the fish lineage, which gave rise to *dpy19l1*, *dpy19l2* and *dpy19l3* [114]. More recent duplication of *dpy19l3* in mammals provided *dpy19l4*. *Dpy19l1* is thought to have undergone ancient gene duplication in its segments of low copy repeats (LCR) and subsequent relocation to give *dpy19l2* [114]. These gene duplications contribute to evolution by acting as molecular substrates for divergence [115], and the process continues to this day, with *dpy19l1* and *dpy19l2* undergoing even more recent gene duplication events leading to the creation of pseudogenes. Two pseudogene exist for *dpy19l1* while *dpy19l2* has four pseudogenes.

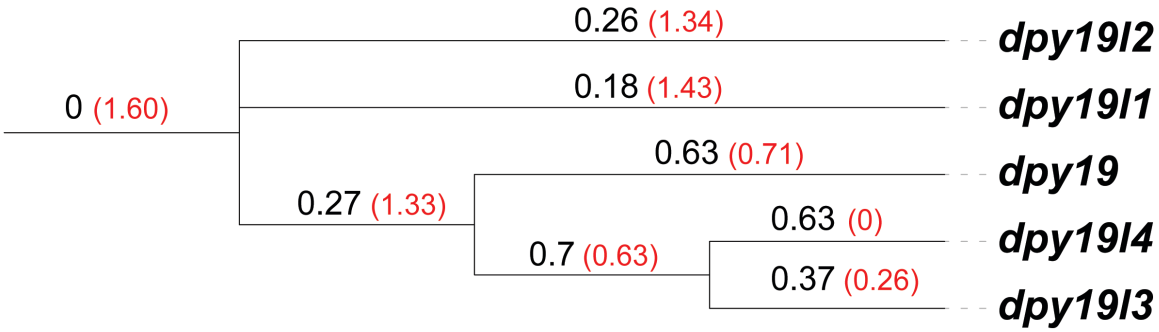


Figure 1. 14: Phylogenetic analysis of the *dpy19* gene family.

The neighbor-joining tree was created using the Jones-Taylor-Thornton matrix for amino acid evolution. *Dpy19* refers to the *C. elegans dpy19* gene. *Dpy19l1-4* are the four human homologues. Numbers in black are raw branch length values. Numbers in red are relative ages.

The number of *dpy19* genes varies significantly across the tree of life (Figure 1.15). As mentioned, mammals have four functional *dpy19* genes, while the Reptilia family only have three. Invertebrates, including nematodes (e.g. *C. elegans*), deuterostomes (e.g. *C. intestinalis*), and insects (e.g. *Drosophila melanogaster*) only have one *dpy19* gene. The apicomplexans (e.g. *Plasmodium* and *Toxoplasma* spp.) also have only one *dpy19* gene, though it was most likely acquired by horizontal gene transfer. Bacteria, fungi and plants do not possess a *dpy19* homologues (Figure 1.15). Clearly, the *dpy19* gene, and thus tryptophan *C*-mannosylation, is an innovation of metazoans. It evolved at the same time as the TSR domain and underwent proliferation in vertebrates at around the same time as the evolution of type-I cytokine receptors, suggesting that the evolution of DPY19 enzymes may be driven by a diversification of substrates.

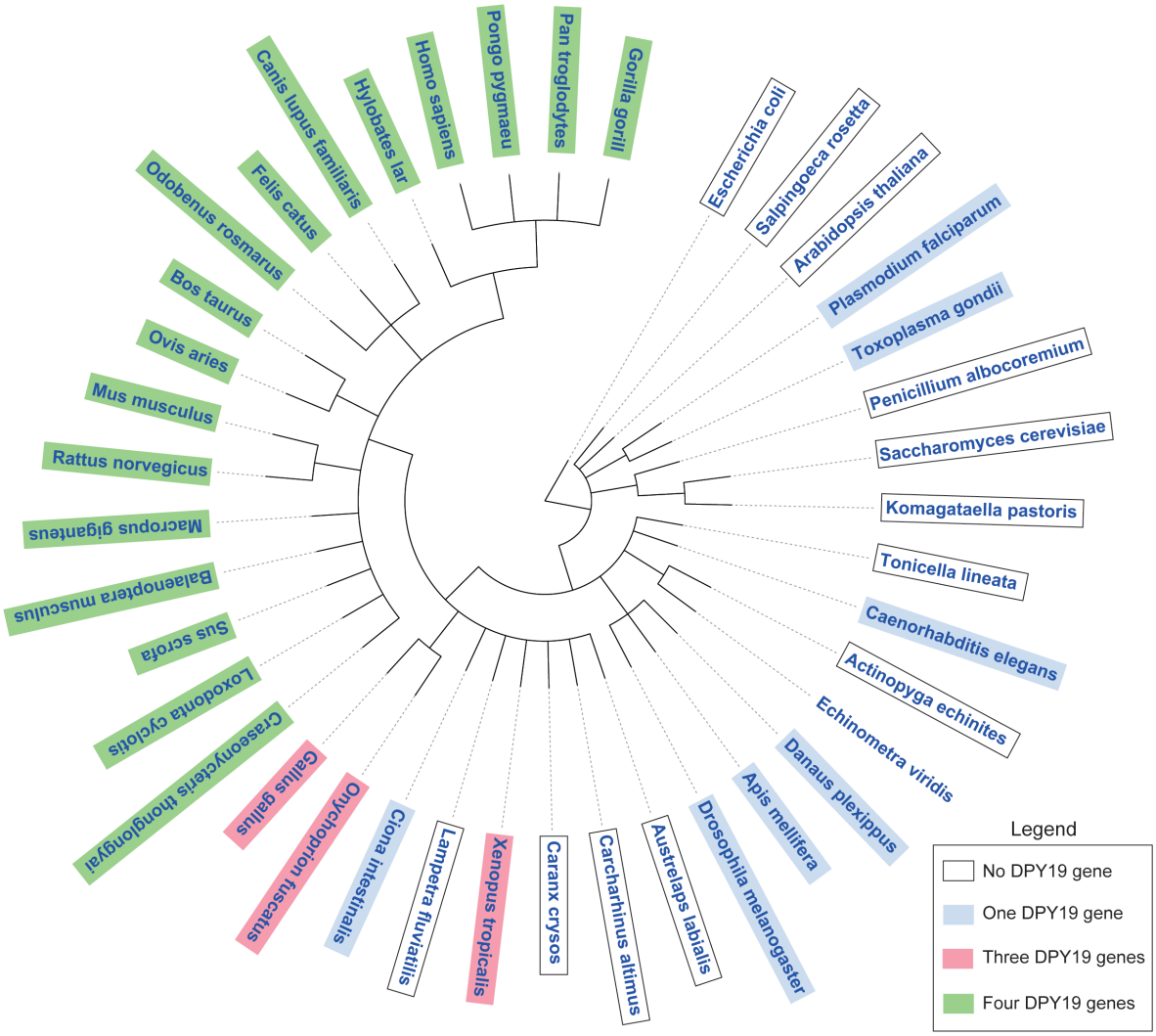


Figure 1. 15: Phylogenetic tree illustrating proliferation of the *dpy19* genes.

The *dpy19* genes all encode proteins with similar structural topology. Each one is predicted to contain ≈10 transmembrane helices, though this number depends somewhat on the topology prediction algorithm used (Figure 1.16). The sequence similarity between *Cedpy19* and the human homologues is not as well conserved as their overall topology. Hydropathy plots reveal three non-transmembrane regions that are assumed to be crucial for catalysis. These regions share far greater sequence similarity between proteins than the transmembrane regions. In all three of these non-membrane regions, human *dpy19l3* and *dpy19l1* have greater homology with *Cedpy19* than the other human homologues.

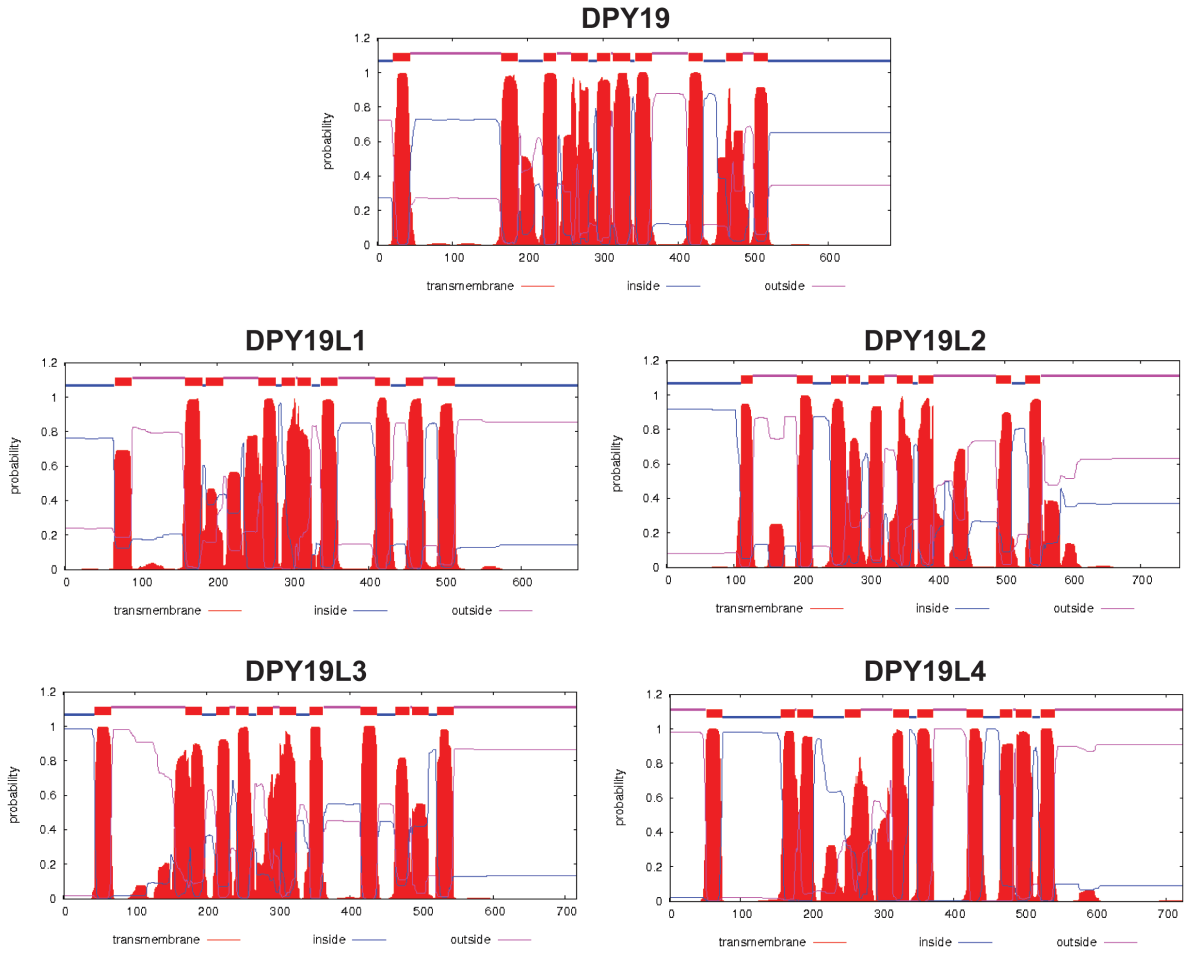


Figure 1. 16: Hydropathy plots for *Cedpy19* and the four human *dpy19* homologues.

Constructed using the TMHHM software package.

1.8 - Mammalian *C*-mannosyltransferases are encoded by *dpy19l1* and *dpy19l3*

In 2015, *dpy19l3*, one of the four mammalian *dpy19* homologues, was identified as a *C*-mannosyltransferase that is active on R-spondin-1 (Rspo1) [71]. In 2017, the Bakker lab discovered that *dpy19l1* was also a *C*-mannosyltransferase [74]. The activity of the gene products from individual *dpy19* genes were examined in isolation through reintroduction to *dpy19l1*, *dpy19l3 dpy19l4* triple knockout CHO cells (*dpy19l2* is only expressed in the testes). Protein substrates, in this case the two TSR domains of mouse UNC5A, were expressed in these cell lines, isolated, and examined by mass spectrometry [74]. The tryptic peptides from TSR-1 (W_1STW$_2$TEW$_3$SVC) and TSR2- (W_1SPW$_2$SKW$_3$SAC) revealed that DPY19L1 had a higher propensity to modify W_1 and W_2, while DPY19L3 had a preference for modifying W_3 [74]. The authors concluded that the preferred peptide substrate for DPY19L1 is WXXW, whilst for DPY19L3 it is WXXC. Given that there is still overlaps in the activities of these

enzymes and numerous examples of non-canonical *C*-mannosylation sites, this conclusion is probably an over simplification. It is also important to keep in mind that this data was obtained from a heterologous system and a single protein substrate at unnatural enzyme and substrate concentrations. As such, it is important not to over-interpret these results but acknowledge that these two *C*-mannosyltransferases, and possibly others, have different and complimentary enzyme activities and as a consequence their biological roles are likely to be different. A brief discussion of the other mammalian *dpy19* genes, which one might assume to be *C*-mannosyltransferases, is warrant.

Dpy19l2 is the most divergent of the *dpy19* genes, is only expressed in the testes and lacks any *C*-mannosyltransferase activity. Mutations in *dpy19l2* are the cause of globozoospermia: the main type of male infertility [116] [117] [118] [119]. DPY19L2 is localised to the inner nuclear membrane facing the acrosomal vesicle [120]. It most likely serves as a structural protein that links the acrosome to the nuclear envelope in spermatozoa. Without DPY19L2, the acrosome does not link to the nucleus, resulting in disrupted vesicular trafficking, lack of nuclear shaping and loss of unbound acrosome [120]. The localisation of DPY19L2, its sequence divergence and dearth of all *C*-mannosylation activity strongly suggest that this protein has evolved a different function altogether.

All attempts to examine DPY19L4 for *C*-mannosyltransferase have provided negative results [74] [71]. The function of this ubiquitously-expressed protein remains unknown. It shares much closer homology with DPY19L1 and DPY19L3 than DPY19L2, so it may be that it is a *C*-mannosyltransferase with a radically different substrate preference, or that it performs a different type of glycosylation. Either way, it remains a protein of great interest and proteomic studies on *dpy19l4* knockout cells lines may be able to provide fresh insights into the role of this protein.

1.9 - Apicomplexan *C*-mannosyltransferases

Recent studies on CRISPR-Cas9-mediated genome-wide genetic knockout screens on *Toxoplasma gondii*, an Apicomplexan parasite, has revealed that a very severe fitness cost is incurred when its *dpy19* homologue is disrupted [121]. Other work has shown that microneme protein 2 (MIC2), which is extremely important for parasite motility and host cell invasion, is *C*-mannosylated by the *T. gondii* parasite on several of its TSR domains [122] [123]. Thus, it

appears likely that the parasite's DPY19 enzyme is important for the trafficking of MIC2 and possibly other parasite virulence factors, though this remains to be proven.

In the malaria parasite, Swearingen *et al*. used mass spectrometry to show that thrombospondin-related adhesive protein (TRAP) from *Plasmodium falciparum* salivary gland sporozoites were *C*-mannosylated on multiple tryptophan residues [124]. Like MIC2, this protein plays an important role in motility and host cell (hepatocyte) invasion [125-127]. Hoppe *et al.* verified that TRAP was *C*-mannosylated by the *P. falciparum* DPY19 enzyme in a heterologous system using protein mass spectrometry [122]. This same system was used to demonstrate that the DPY19 enzymes from these parasites did not modify the WAKW peptide but did recognise the WAEWGEC peptide [122]. It remains unclear as to whether the peptide's size, charge or sequence was the defining difference here, though it illustrates once again that the canonical WXXW *C*-mannosylation motif is an oversimplification that must evolve as we learn more about the diversity of the DPY19 enzymes.

1.10 - Summary

This chapter has highlighted the key advancements in the tryptophan *C*-mannosylation field since its discovery in 1994. For a long time, the field has taken an *ad hoc* approach to identifying this modification and sought to explore the role of the modification through imperfect mutagenesis experiments. The recent discovery of genes encoding *C*-mannosyltransferases and a growing list of non-canonical *C*-mannosylation sites has changed the way we think about tryptophan *C*-mannosylation and provides a plethora of opportunities for the discovery of new glycobiology. This thesis outlines the experiments, results and conclusions I have made during my PhD to advance our understanding of this unusual protein modification.

1.11 - Aims

1. Create a simple expression system for the production of proteins with and without tryptophan *C*-mannosylation by engineering a *C*-mannosyltransferase into the yeast *Pichia pastoris*.
2. Characterisation of the impact of tryptophan *C*-mannosylation on protein stability and function.

3. Obtain insights into the mechanism of *C*-mannosylation by:
    a) Identifying conserved and functionally essential residues in the tryptophan *C*-mannosyltransferases.
    b) Probing the substrate preference of tryptophan *C*-mannosyltransferases.
4. Produce and validate monoclonal antibodies against C^2-α-D-mannosyl-tryptophan to enable the simple identification of *C*-mannosylated proteins using immunochemistry techniques.
5. Generation of chemical inhibitors of *C*-mannosylation for probing the function of *C*-mannosylation in biological systems.
6. Knockout of the mammalian *C*-mannosyltransferase genes, *dpy19l1* and *dpy19l3*, in a human cell line, and probe these systems using surface proteomics to provide characterisation of the loss of *C*-mannosylation in a biological system.

Chapter 2

Glycoengineering a tryptophan mannosylation pathway into *Pichia pastoris*

2.1 - Introduction

To study the impact of tryptophan *C*-mannosylation on protein fold and function we required a simple protein expression system to facilitate the production of samples with and without tryptophan mannosylation. The DPY19 *C*-mannosyltransferase enzymes utilise mannose-phosphate-dolichol (Dol-P-Man) as a glycosyl donor. This molecule is only found in the eukaryotic ER membrane, which precluded the use of bacterial expression systems. Most metazoans have several homologues of DPY19 but simple eukaryotes like yeast have none (Table 2.1). Enzyme activity assays on yeast microsomes have confirmed their inability to perform tryptophan *C*-mannosylation [128]. Collectively, this suggested that we needed to either disrupt the multiple DPY19 genes in mammalian or insect cell lines or introduce DPY19 genes into simple eukaryotes like yeast to obtain a common expression platform to make proteins with and without the modification. We elected to take the former approach and engineer the yeast *Pichia pastoris* because it is very simple to genetically manipulate, provides very stable cell lines, produces copious amounts of Dol-P-Man substrate for the DPY19 enzymes, and is a high-yielding protein production platform. Furthermore, yeast is not accustomed to *C*-mannosylation and will not have machinery that recognise C^2-α-D-mannosyl-tryptophan such as lectins or chaperons that will help proteins fold. Therefore, any effects seen in proteins produced in the engineered yeast strain is due directly to the C^2-α-D-mannosyl-tryptophan attached to the proteins.

Table 2. 1: List of organisms with the DPY19 gene, enzyme activity and Dol-P-Man

Organism	DPY19 Genes	Enzyme activity	Dol-P-Mannose
Pan troglodytes	4	Yes	Yes
Bos taurus	4	Yes	Yes
Mus musculus	4	Yes	Yes
Homo sapiens	4	Yes	Yes
Gallus gallus	3	Yes	Yes
Xenopus tropicalis	3	Yes	Yes
Drosophila melanogaster	1	?	Yes
Danaus plexippus	1	No	Yes
Caenorhabditis elegans	1	Yes	Yes
Toxoplasma gondii	1	Yes	Yes
Plasmodium falciparum	1	Yes	Yes
Pichia pastoris	0	No	Yes
Arabidopsis thaliana	0	No	Yes
Escherichia coli	0	No	No

Our approach was to produce *P. pastoris* cell lines that constitutively express DPY19 homologues (Table 2.2). We chose to use the pGAPz vector system for this because it provides constitutive expression of DPY19, under the powerful GAP promoter and protein would be expressing intracellularly and selectable using zeocin (Table 2.2). Various inducible expression constructs could then be integrated into these cell lines, and the parental cell line lacking DPY19, in order to produce proteins with and without tryptophan *C*-mannosylation. We chose the pPIC9k system for this because it provides inducible expression under the AOX1 promoter and provides selection using HIS4. This system provides secreted expression of target proteins and would mean that our reporter proteins can travel through the secretary pathway and come in contact with DPY19 (Table 2.2). The DPY19-expressing cell lines also provide a means to study the peptide substrate preferences and DPY19 mechanism *in vitro*. Such a simple protein production platform might also be useful in the biotechnology sector for the manufacture of vaccines with native glycosylation patterns, including those based on the Ebola virus soluble glycoprotein (EBOV sGP) [129] and the malaria parasite's thrombospondin-related adhesive protein (TRAP), which is the antigen in a candidate vaccine for malaria [130]. Native EBOV sGP and TRAP are *C*-mannosylated and correct glycosylation is likely important for maximising vaccine efficacy and this will require engineering of *C*-mannosylation pathways into the yeast expression hosts used to make these vaccines [111] [124].

Table 2. 2: Summary of the some of the commercial vector systems available for *P. pastoris* protein expression

Vector	Promotor	Strain	Bacterial selection	Pichia selection	Expression	Regulation	Protein
pFLD1	FLD1	X-33, GS115, KM71H	Amp & Zeo	Zeocin	Inducible	meOH	Intracellular
pPIC6	AOX1	X-33	Blasticidin	Blasticidin	Inducible	meOH	Secreted
pPICZα	AOX1	X-33,KM71H	Zeocin	Zeocin	Inducible	meOH	Secreted
pGAPZ	GAP	GS115, X-33, KM71	Zeocin	Zeocin	Constitutive	glucose	Intra & sec
pPIC3.5K	AOX1	GS115, KM71	Amp & Gent	Gen & HIS4	Inducible	meOH	Intracellular
pPIC9K	AOX1	GS115, KM71	Amp & Gent	Gen & HIS4	Inducible	meOH	Secreted

There is precedence for engineering yeast glycosylation pathways to provide biologics with native glycosylation. Most notably, this has been accomplished with the *N*- and *O*-glycosylation pathways (Figure 2.1A, B). There are many *N*- and *O*-glycan structures, which can be complex, and the examples given below are not represented for all of the *N*- and *O*-glycans that are seen in humans. *P. pastoris* is known to hypermannosylate *N*-glycans on proteins, and the enzyme responsible for this is the mannosyltransferase (OCH1) gene [131]. The first attempt to humanise the *N*-glycosylation pathway in *P. pastoris* was done in 2001 when a 1,2-α-D-mannosidase was knocked into *P. pastoris* [132]. This enzyme trims the 1,2-α-D-mannose that elongates the hypermannosylated *N*-glycan but this approach was only 85% efficient, giving heterogeneous *N*-glycans [132].

Further work was done to produce proteins with human-like non-sialylated hybrid-type *N*-linked oligosaccharide structure by knocking-out the OCH1 in the 1,2-α-D-mannosidase over expressing cell line along with the knock-in of both *N*-acetylglucosaminyltransferase I and beta-1,4-galactosyltransferase [133]. Meanwhile, another research group, at the same time, was able to produce recombinant proteins with human-like complex *N*-linked oligosaccharide structures [134]. This was done using the YJN201 strain of *P. pastoris*, which lacks OCH1 activity, with knock-in of α-1,2-mannosidase, and β-1,2-*N*-acetylglucosaminyltransferase I (GnTI) and UDP-GlcNAc transporter [135]. They further modified the strain by knocking-in mannosidase II and β-1,2-*N*-acetylglucosaminyltransferase II (GnTII) and were able to produce proteins with *N*-$GlcNAc_2Man_3GlcNAc_2$ complex human *N*-glycan [134]. In 2009 a nature

protocols paper was published for the production of different complex and hybrid *N*-glycans in *P. pastoris* using a technology called Glycoswitch [136]. The Glycoswitch technology takes advantage of using a step-wise introduction of different glycosidases and glycosyltransferases in different combinations to reach desired *N*-glycan structures [136].

The next feat of engineering was to produce complex terminally sialylated *N*-glycan structures. This was challenging because the sialylation pathway, the final step in human *N*-glycosylation, does not exist in yeast [137]. Complex *N*-glycans with terminal sialylation are important for glycoproteins, although not for antibodies, because the sialic acid increases their half-life in the blood stream and their potency [137]. A strain of *P. pastoris* called GS.5 which produces terminally galactosylated biantennary glycans of the complex type $Gal_2GlcNAc_2Man_3GlcNAc_2$ was used for engineering the new strain [138]. Five different enzymes were knocked in to generate the new strain and recapitulate the entire sialylation pathway into yeast [137]. They were able to produce recombinant human erythropoietin (EPO) that was comparable to biologically active EPO produced in mammalian systems [137].

Much like *N*-glycans in yeast, *O*-linked glycans are also hypermannosylated in yeast and have been subject to humanisation. Protein mannosyltransferases (PMT), of which there are 5 homologues in *P. pastoris*, extend the chain of mannose residues on *O*-linked glycans and PMT1 and PMT2 have been knocked out in *P. pastoris* along with the use of PMT inhibitors greatly reduce the degree of hypermannosylation [139]. However, this approach alone does not completely remove the mannose residues, and often α-1,2-mannosidase and lysosomal mannosidase are knocked into these strains to remove the mannose residues or *in vitro* treatment of glycoproteins with recombinant mannosidases is performed [140] [141].

An example of further humanising *O*-linked glycosylation in yeast is the engineering of glycans to containing Galβ1,3GalNAc1, which is a mucin type glycosylation, engineered in *S. cerevisiae* [142]. Four genes were introduced into the yeast as well as using PMT inhibitors to block endogenous *O*-linked hypermannosylation. Although this was not done in *P. pastoris*, it provides a proof-of-principle for it to be done in *P. pastoris*. Another example is the engineering of the α-dystroglycan *O*-linked glycosylation pathway into *P. pastoris*. An α-1,2-mannosidase was introduced into the cell to trim the α-1,2-mannose chain, followed by the introduction of protein-*O*-linked-mannose β-1,2-*N*-acetylglucosaminyltransferase 1 which

capped the glycan with *N*-acetylglucosamine and then the sialylation pathway was introduced to add a sialic acid to the glycan [141].

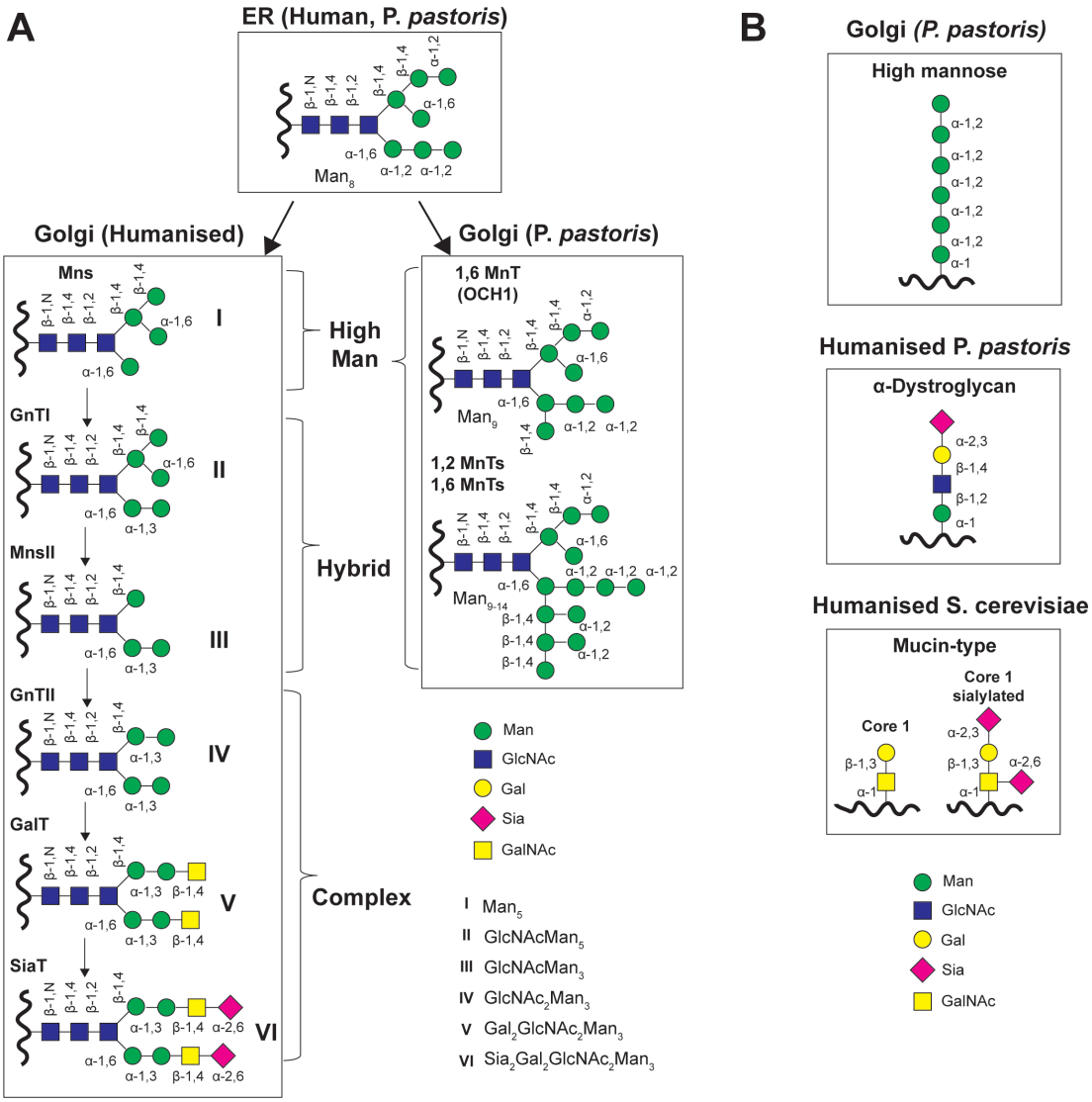


Figure 2. 1: Humanisation of the *N*- and *O*-glycosylation pathways in yeast

a) Humanisation of the *N*-glycosylation pathway in *P. pastoris*

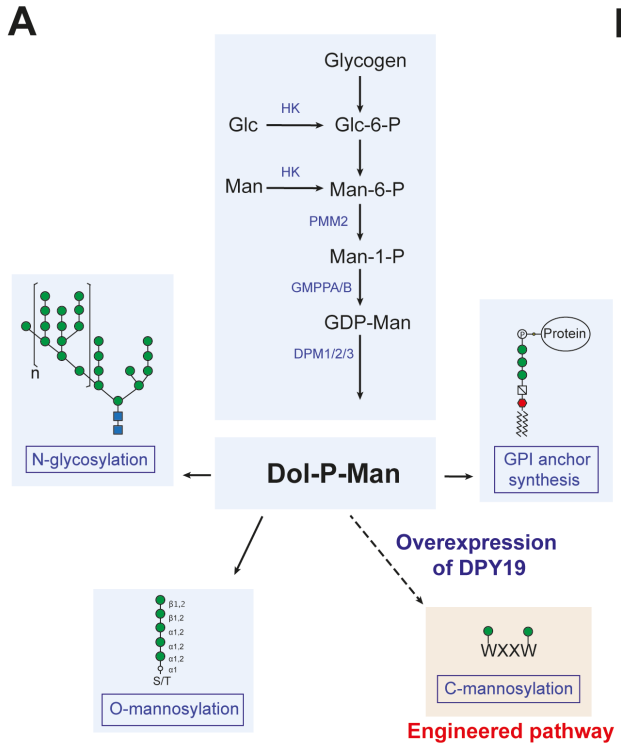
b) Humanisation of the *O*-glycosylation pathway in *P. pastoris* and *S. cerevisiae*

With so much progress on the humanisation of *P. pastoris N*- and *O*-glycosylation pathways, it was now time to compliment these efforts by engineering a mammalian-like *C*-mannosylation pathway into this versatile yeast.

2.2 - Results

2.2.1 - Probing the suitability of *P. pastoris* as a host for *C*-mannosyltransferases

We chose *P. pastoris* as a host for DPY19 enzymes because the yeast ER possesses a large reservoir of Dol-P-Man that is ordinarily utilised for the synthesis of high-mannose *N*-glycans, *O*-mannosylation and GPI anchor synthesis (Figure 2.2A) [143]. However, before introducing any DPY19 homologues in the yeast, we wanted to ensure that there were no obvious toxicity issues that could arise from *C*-mannosylation of essential yeast proteins passing through the secretory pathways. A BLASTp search revealed that 345 *P. pastoris* proteins possessed the WXXW consensus motif for *C*-mannosylation and that only 27 of these proteins pass through the secretory pathway (Figure 2.2B). None of these proteins were known to be essential and so we hypothesised that introduction of the DPY19 enzyme would not have any significant impact on the fitness of the yeast.



B

Ascension	Protein
XP_002491059.1	Hypothetical protein
XP_002493381.1	RNA polymerase II mediator subunit
XP_002493168.1	Beta-1,6 glucan biosynthesis protein
XP_002489455.1	Beta-1,6 glucan biosynthesis protein
XP_002493349.1	Serine-threonine kinase
XP_002493332.1	Hypothetical protein
XP_002494060.1	Hypothetical protein
XP_002491202.1	Hypothetical protein
XP_002489477.1	Hypothetical protein
XP_002493648.1	Ferric reductase
XP_002493224.1	Ferric and cupric reductase
XP_002491319.1	Transmembrane protein
XP_002491562.1	Hypothetical protein
XP_002490607.1	Hypothetical protein
XP_002492388.1	Glucosidase II beta subunit
XP_002491707.1	Mannan polymerase I subunit
XP_002490358.1	Putative mannosidase
XP_002489792.1	Putative mannosidase
XP_002494380.1	Vacuolar protein sorting
XP_002494132.1	Exo-1,3-beta-glucanase
XP_002489860.1	Hypothetical protein
XP_002491855.1	Hypothetical protein
XP_002489746.1	Chitin deacetylase
XP_002492097.1	Type I membrane protein
XP_002492925.1	Hypothetical protein
XP_002493445.1	Hypothetical protein
XP_002492833.1	Hypothetical protein

Figure 2. 2: Engineering a glycosylation pathway into *P. pastoris*.

a) Diagram illustrating the strategy for humanization of the yeast ER glycosylation pathway. Man: D-mannose; Glc: D-glucose; HK: hexokinase; MPI, phosphomannose isomerase; PMM2: phosphomannomutase; GMPPA/B: GDP-mannose

pyrophosphorylase (A/B); PGI: phosphoglucose isomerase; KPS: KDN-9-phosphate synthase; Man-6-P: mannose-6-phosphate; Fru-6-P: fructose-6- phosphate; Glc-6-P: glucose-6-phosphate; GDP-Man: GDP-mannose; Dol-P-Man: dolichol phosphate mannose; DPM1/2/3: Dol-P-Man synthase.

b) Table of yeast proteins with the WXXW motif and that pass through the secretory pathway.

2.2.2 - Engineering an active tryptophan *C*-mannosyltransferase in *P. pastoris GS115*

We attempted to express three different tryptophan *C*-mannosyltransferases into *P. pastoris GS115*, including *Caenorhabditis elegans* DPY19 and the two human enzymes DPY19L1 and DPY19L3. In each case, genes encoding these enzymes with a C-terminal 3×FLAG tag were codon optimised for *P. pastoris*, synthesised and cloned into the pGAPZ (Invitrogen) vector under the strong constitutive glyceraldehyde-3-phosphate dehydrogenase (GAP) promoter (Figure 2.3A). The plasmid was linearised within the GAP promoter using the *AvrII* restriction enzyme and the purified linear DNA electroporated into *GS115* cells. Integrants were selected for using zeocin. Two dozen individual colonies from each transformation were cultured and used to prepare genomic DNA. PCR confirmed integration of the pGAPz-DPY19 cassettes into the GAP locus (Figure 3B). Western blot analyses using an antibody to the FLAG peptide revealed no evidence of FLAG-tagged peptides for the human DPY19L1 or DPY19L3 (Figure 2.3C), despite having genomic integration verified through PCR (data not shown). However a strong band around 55 kDa in size was detected for cells harbouring *C. elegans* DPY19 (Figure 2.3D). This band migrated significantly faster than expected for *Ce*DPY19-$FLAG_3$, which has a molecular weight of 83 kDa. Since the C-terminal FLAG-tag was detected by the antibody, the protein is either truncated at the N-terminus or, more likely, simply migrates faster-than-usual by SDS-PAGE due to the high percentage of hydrophobic amino acids in this type-IV integral membrane protein [144]. Hydrophobic membrane proteins bind more SDS than globular proteins and so carry a higher charge to size ratio, resulting in faster migration through the acrylamide gel. This explanation is consistent with the considerable sample preparation optimisation that was needed for to prevent the DPY19 precipitating from sample buffer: lysis had to be performed in 10% SDS at 37 $^{\circ}$C for 10 minutes immediately prior to running the SDS-PAGE to prevent the formation of higher molecular weight aggregates.

To test if the *Ce*DPY19 was localised to the microsomal compartments, yeast cells were fractionated into their subcellular components by ultracentrifugation and another western blot was performed on each fraction. The microsomal fractions, most likely the ER, contained active *Ce*DPY19 (Figure 2.3E). To quantify the amount of *Ce*DPY19 in the yeast, a titration was performed by western blot using maltose-binding protein (MBP) fused to the 3×FLAG as a standard (Figure 2.3F). Densitometry analysis indicated that there were around 4×10^5 copies of *Ce*DPY19 per cell, which is an order of magnitude higher than for the oligosaccharyltransferase components OSTA and OSTB, which perform *N*-glycosylation within the ER. This very large number is of the same order of magnitude as for GAP, which was the promotor used for DPY19 in this system (Figure 2.3G) [145].

Growth curves for the parental *GS115* and the DPY19 cell lines demonstrated that there was no obvious fitness cost resulting from the overexpression of DPY19 (Figure 2.3H, I). An SDS-PAGE for the supernatant of the *GS115* and DPY19 strains grown in YPD and BMMY showed no obvious differences in protein secretion profiles (Figure 2.3J). Given the lack of an obvious phenotype upon introduction of DPY19, we did not investigate changes in the proteome or metabolome any further.

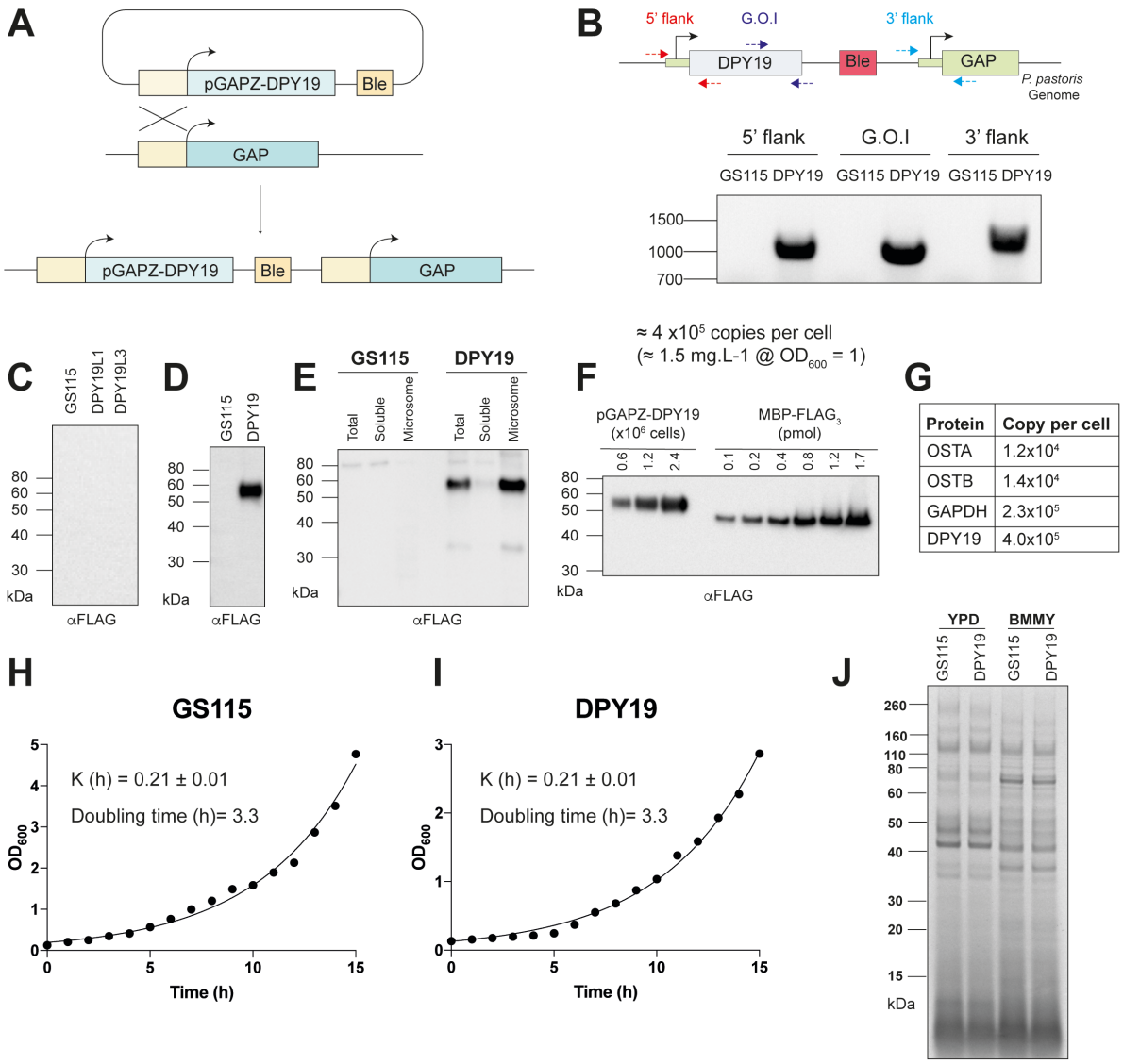


Figure 2. 3: Expression of tryptophan *C*-mannosyltransferases strain in *P. pastoris GS115*

a) A cassette encoding DPY19 with a zeocin selectable marker was integrated into the *P. pastoris* genome by homologous recombination at the GAP promoter.

b) Integration of the cassette encoding DPY19 was confirmed using PCR

c) A representative western blot (α-FLAG)) of total lysates from the *GS115*, DPY19L1 and DPY19L3 strains.

d) A representative western blot (α-FLAG) of total lysate from the *GS115* and DPY19 strains.

e) A representative western blot (α-FLAG) of sub-cellular fractions of cells from the *GS115* and DPY19 strains. Total: total lysate; Soluble: cytosolic fraction; Microsome: microsomal fraction containing ER, Golgi and Mitochondria.

f) A representative western blot (α-FLAG) quantifying DPY19 copy number per cell using MBP-FLAG as a calibration standard. Densitometry was performed to estimate the copy number of DPY19 protein per cell.
g) Table showing the copy number per cell of OSTA and OSTB, subunits making up the oligosaccharyltransferase, which perform *N*-glycosylation.
h) Growth curve for *GS115* demonstrating its exponential growth curve. OD_{600} measurement were taken of *GS115* (n=3) every 1 h for 15 h. An exponential growth curve was fit using Prism to obtain the rate constant k, from which doubling times were calculated.
i) Growth curve for DPY19 demonstrating that there is no growth defect for cells harbouring DPY19. OD_{600} measurement were taken of DPY19 (n=3) every 1 h for 15 h. An exponential growth curve was fit using Prism to obtain the rate constant k, from which doubling times were calculated.
j) SDS-PAGE of yeast supernatants from *GS115* and DPY19 strains grown in both YPD and BMMY media. SDS-PAGE gel is stained with Coomassie.

2.2.3 – *Ce*DPY19 in *P. pastoris* microsomes is catalytically competent

An *in vitro C*-mannosylation radioassay was adapted for use with yeast microsomes to investigate if the recombinant *Ce*DPY19 was active [14]. This radioassay utilises endogenous dolichol kinase and ATP to convert dolichol in the microsomal membranes to dolichol phosphate. Endogenous dolichol-phosphate-mannose synthase then utilises GDP to convert dolichol phosphate into Dol-P-Man. Use of GDP-[^{3}H]-mannose in this assay results in the generation of radiolabelled Dol-P-[^{3}H]-Man. This radiolabelled mannose is then transferred to tryptophan residues in the WXXW acceptor peptide. Purification of the peptide from the reaction mixture and scintillation counting enables one to detect tryptophan *C*-mannosyltransferase activity (Figure 2.4A).

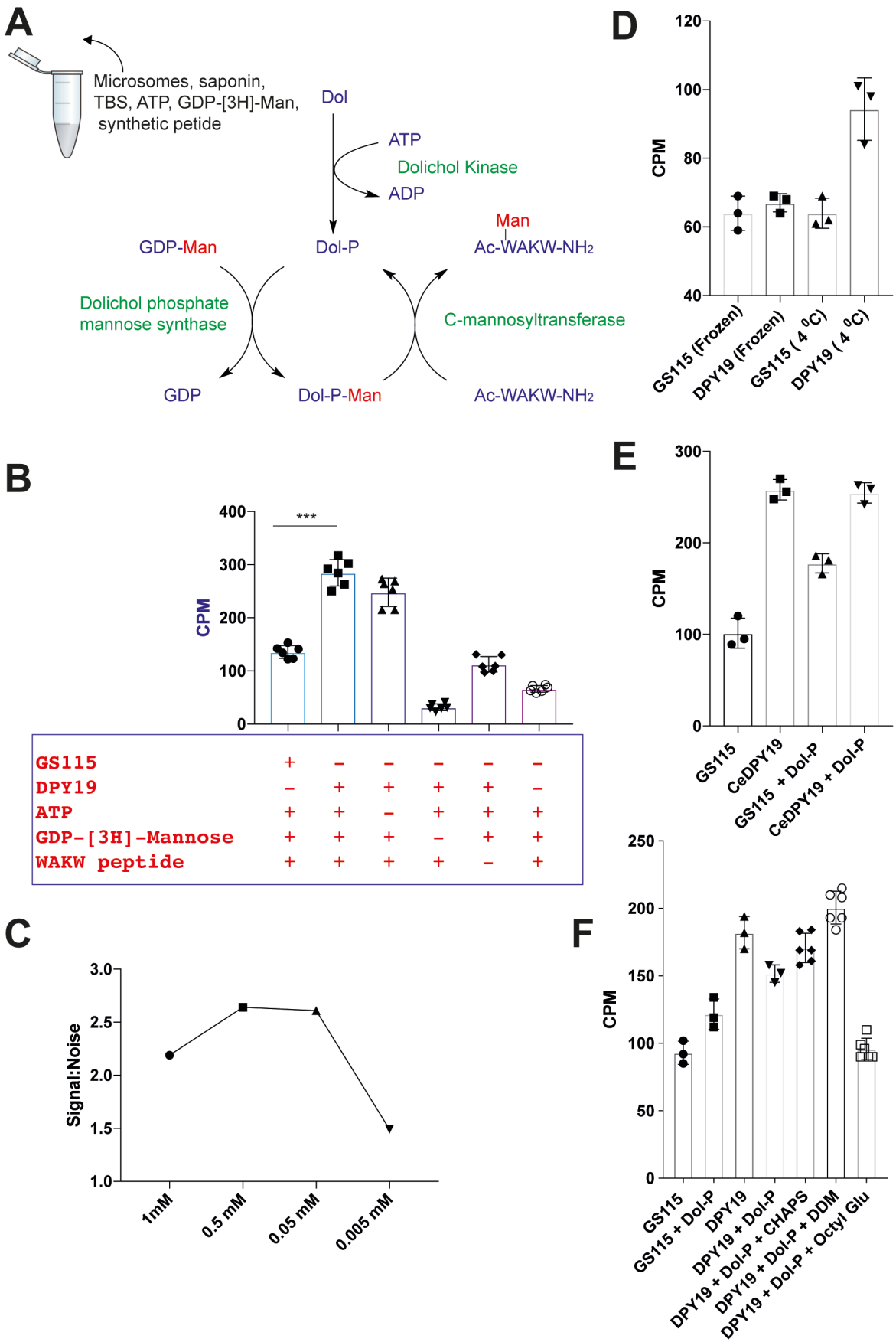

A
Microsomes, saponin,
TBS, ATP, GDP-[3H]-Man,
synthetic petide
Dol
ATP
Dolichol Kinase
ADP
GDP-Man
Dol-P
Man
Ac-WAKW-NH2
Dolichol phosphate
mannose synthase
C-mannosyltransferase
GDP
Dol-P-Man
Ac-WAKW-NH2
D
CPM
120
100
80
60
40
GS115 (Frozen)
DPY19 (Frozen)
GS115 (4 °C)
DPY19 (4 °C)
B
CPM
400
300
200
100
0
***
GS115 + – – – – –
DPY19 – + + + + –
ATP + + – + + +
GDP-[3H]-Mannose + + + – + +
WAKW peptide + + + + – +
E
CPM
300
200
100
0
GS115
CeDPY19
GS115 + Dol-P
CeDPY19 + Dol-P
C
Signal:Noise
3.0
2.5
2.0
1.5
1.0
1mM
0.5 mM
0.05 mM
0.005 mM
F
CPM
250
200
150
100
50
0
GS115
GS115 + Dol-P
DPY19
DPY19 + Dol-P
DPY19 + Dol-P + CHAPS
DPY19 + Dol-P + DDM
DPY19 + Dol-P + Octyl Glu

Figure 2. 4: DPY19 is active in yeast and can transfer radiolabelled mannose onto acceptor peptides.

a) Diagram showing the *in vitro C*-mannosylation reaction components and pathway. TBS; Tris-buffered saline; ATP: Adenosine triphosphate; ADP: Adenosine diphosphate; Dol: dolichol; Dol-P: dolichol phosphate; Dol-P-Man: dolichol phosphate mannose; GDP: guanosine diphosphate; GDP-Man: GDP-mannose.
b) Synthetic WAKW acceptor tetrapeptide titration curve. Peptide concentrations of 1 mM (n=3), 0.5 mM (n=3), 0.05 mM (n=3) and 0.005 mM (n=3) were assayed. The signal to noise-ratio was measured for each concentration against reactions performed with *GS115* (n=3).
c) *In vitro C*-mannosylation assay using microsomes from the *GS115* and DPY19 strains. All results are mean values ± SD. *** $p<0.005$.
d) *In vitro C*-mannosylation assay using microsomes that were stored at -80 °C and at 4 °C overnight. All results are mean values ± SD.
e) *In vitro C*-mannosylation assay using Dol-P. Dol-P was at 200 μM. All results are mean values ± SD.
f) *In vitro C*-mannosylation assay using Dol-P with different detergents. Dol-P was at 200 μM. DDM: *n*-Dodecyl β-D-maltoside; OGT: octyl glucoside. All results are mean values ± SD.

Microsome fractions containing DPY19 enzyme isolated from the DPY19 yeast strain showed significant activity as compared to the parental *GS115* strain (Figure 2.4B). Omission of ATP did not dramatically impact mannosylation, suggesting that the dolichol phosphate concentration was not rate limiting for this coupled enzyme assay (Figure 2.4B). Omitting the acceptor peptide, the radiolabelled GDP-Man or microsomes from the experiment abrogated signal, as expected (Figure 2.4B). A titration of the acceptor peptide was performed to optimise the conditions of the DPY19 activity (Figure 2.4C). Optimal signal-to-noise was obtained for a peptide concentration between 50 and 500 μM. Higher concentrations increased the background signal and concentrations as low as 5 μM gave very little signal at all. To test the stability of DPY19 under different conditions, microsomes that were flash frozen at -80 °C or stored overnight at 4 °C were used in the assay. The results indicated that one freeze/thaw cycle or overnight storage at 4 °C was sufficient to inactivate the microsomes, which could be due to denaturation of DPY19 or one of the other enzymes involved in this coupled assay (Figure 2.4D). We sought to the improve the signal-to-noise ratio of the assay and circumvent any

issues associated with the stability of dolichol kinase by the addition of dolichol-phosphate (Dol-P) to the reactions, as has been reported in mammalian systems [14]. However, we saw no improvement in signal from the assay or stability in microsome activity (Figure 2.4E) but did observe some precipitate form in these reactions. We assessed the impact of the detergents CHAPS, *n*-Dodecyl β-D-maltoside (DDM) and octyl glucoside (OGT) on the assay in the presence of Dol-P. While these detergents reduced visible precipitation, CHAPS and DDM did not significantly improve activity and OGT completely ablated activity (Figure 2.4F). Clearly, OGT may inhibit or denature DPY19 or other enzymes in the coupled assay. Collectively, the limited stability of enzyme activity in this microsomal assay coupled with the arduous process of preparing yeast microsomes dramatically diminished the utility of this assay.

2.2.4 – Expression of *C*-mannosylated proteins in the engineered yeast strain

Having established that *Ce*DPY19 was active in yeast microsomes, the next step was to determine if the enzyme could commandeer intracellular Dol-P-Man to modify heterologous proteins passing through the yeast's secretory pathway. To test this, a number of constructs encoding *C*-mannosylated proteins were prepared using the pPIC9k plasmid. This plasmid makes use of the strong inducible *AOX1* promotor, which regulates the expression of alcohol oxidase 1 in *P. pastoris* to enable this methylotrophic yeast to utilise methanol and other alcohols as a source of carbon [146]. This is a very strong inducible promotor that is commonly used for recombinant protein expression in *P. pastoris* [147]. These plasmids were to be transformed into the *GS115* parental *P. pastoris* strain and our engineered line that constitutively expresses *Ce*DPY19. Integrants would then be selected for using histidine dropout media, since both of these cell lines have loss-of-function mutations in the *his4* locus, which encodes histidinol dehydrogenase, and the pPIC9k plasmid contains a functional copy of this locus (Figure 2.5A).

Genes encoding the ectodomain of human type-I cytokine receptors, human and *C. elegans* TSRs, the EBOV sGP and RNaseII were codon-optimised for *P. pastoris*, synthesized and cloned into the multiple cloning site of a pPIC9k plasmid that was modified to remove the STE13 protease site but retain the α factor pre/pro signal peptide [148, 149] (Figure 2.5B). The extracellular domain of the human type I cytokine receptors IL21R, EPOR and PLR were fused to a C-terminal hexahistidine tag and the human cytokine IL12B was fused to an N-terminal hexahistidine tag (Figure 2.5B). The TSR domains of human spondin-2 (mindin) and *C.*

elegans UNC5, as well as human RNaseII, were fused to an N-terminal FLAG, hexahistidine and SUMO tag as well as a C-terminal Strep-II tag (Figure 2.5B). EBOV sGP was fused to a C-terminal hexahistidine tag (Figure 2.5B).

Clones for each construct in the *GS115* and DPY19 strains were selected for growth on histidine dropout media and for their ability to grow on methanol. Integration of the expression construct was confirmed by PCR and individual clones were chosen for scale-up based on the results of small-scale protein test expressions for each clone. Protein was obtained in at least one of the yeast strains for all constructs except EBOV sGP. Protein production was scaled-up for the other seven constructs in both yeast strains. Briefly, scale-up involved growth of the yeast in BMGY media and induction in BMMY media, with supernatants being harvested after 48 h (Figure 2.5C), concentrated and dialysed prior to immobilised metal ion affinity chromatography (IMAC), size exclusion chromatography (SEC) and ion exchange chromatography to obtain pure recombinant protein, as determined by SDS-PAGE (Figure 2.6A).

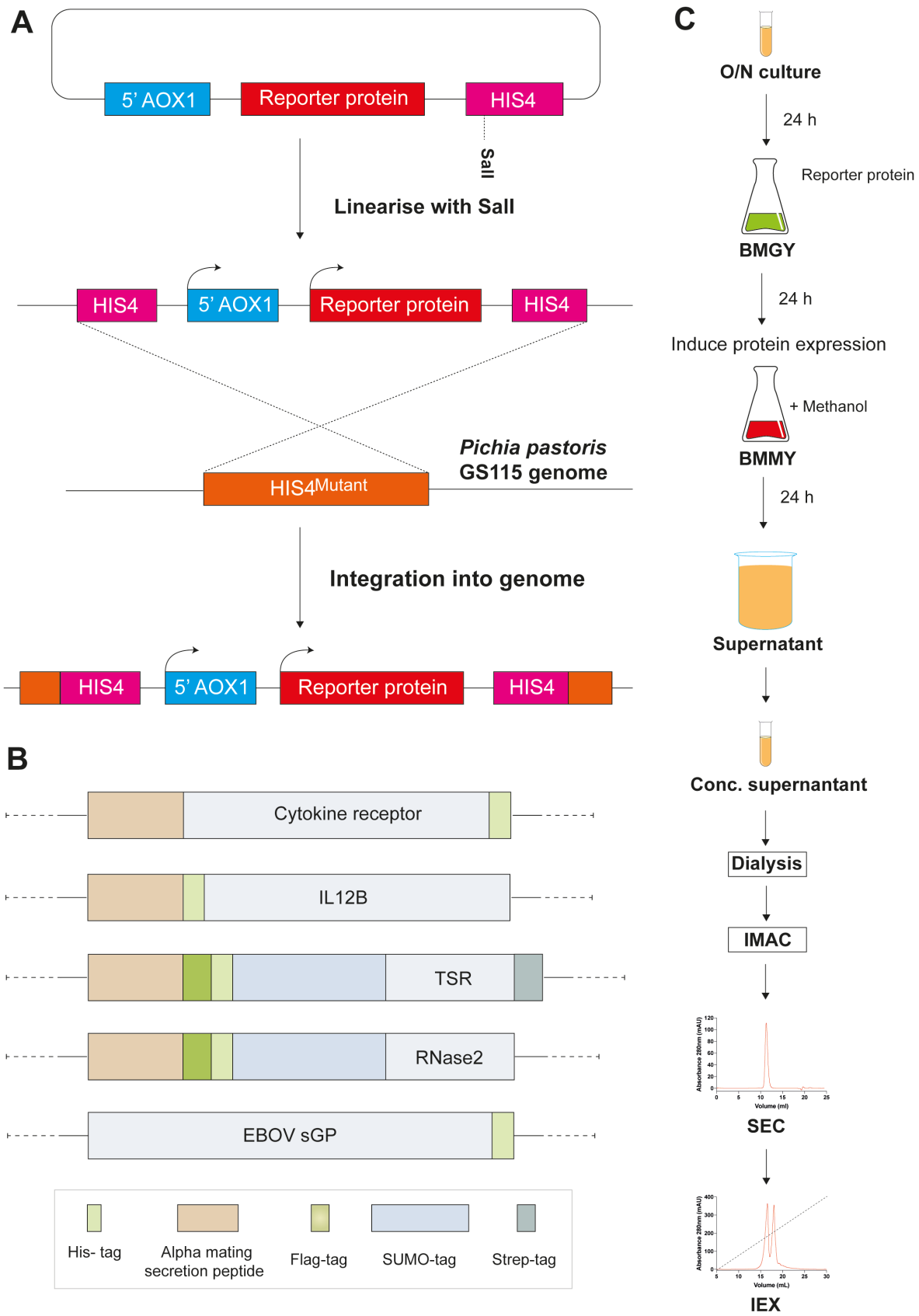


Figure 2. 5: Protein expression and purification strategy in the *GS115* and DPY19 strains.

a) Integration of the pPIC9K cassette into the *P. pastoris* genome at the *AOX1* promoter with *his4* complementation.
b) Construct design for reporter proteins produced in the *GS115* and DPY19 strains. Cytokine receptors: IL21R, EPOR and PLR; TSRs: UNC5 and mindin.
c) The protein expression and purification work-flow.

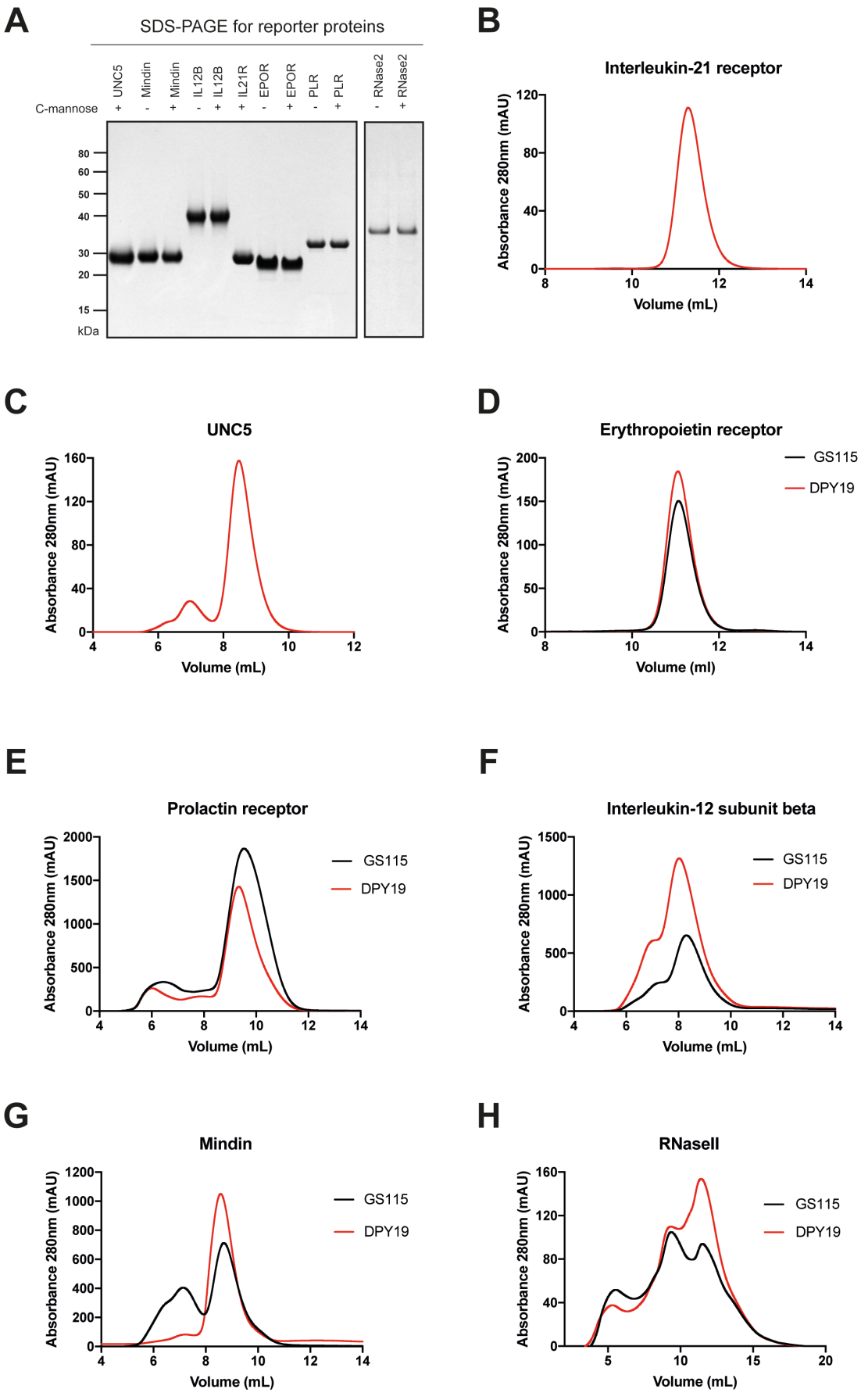

A
SDS-PAGE for reporter proteins
C-mannose
+ UNC5
- Mindin
+ Mindin
- IL12B
+ IL12B
+ IL21R
- EPOR
+ EPOR
- PLR
+ PLR
- RNase2
+ RNase2
80
60
50
40
30
20
15
kDa
B
Interleukin-21 receptor
Absorbance 280nm (mAU)
Volume (mL)
C
UNC5
Absorbance 280nm (mAU)
Volume (mL)
D
Erythropoietin receptor
GS115
DPY19
Absorbance 280nm (mAU)
Volume (ml)
E
Prolactin receptor
GS115
DPY19
Absorbance 280nm (mAU)
Volume (mL)
F
Interleukin-12 subunit beta
GS115
DPY19
Absorbance 280nm (mAU)
Volume (mL)
G
Mindin
GS115
DPY19
Absorbance 280nm (mAU)
Volume (mL)
H
RNaseII
GS115
DPY19
Absorbance 280nm (mAU)
Volume (mL)

Figure 2. 6: Purification of reporter proteins produced in the *GS115* and DPY19 strains.

a) SDS-PAGE of reporter proteins purified from the *GS115* and DPY19 strains. 3 µg of protein was combined with 4X SDS loading buffer and heated at 95 °C for 2 min and then loaded onto a SDS-PAGE gel.
b) Size-exclusion chromatogram for IL21R-DPY19 using a s75 10/300 column.
c) Size-exclusion chromatogram for UNC5-DPY19 using a s75 10/300 column.
d) Size-exclusion chromatogram for EPOR-GS115 and EPOR-DPY19 using a s75 10/300 column.
e) Size-exclusion chromatogram for PLR-GS115 and PLR-DPY19 using a s75 10/300 column.
f) Size-exclusion chromatogram for IL12B-GS115 and IL12B-DPY19 using a s75 10/300 column.
g) Size-exclusion chromatogram for mindin-GS115 and mindin-DPY19 using a s75 10/300 column.
h) Size-exclusion chromatogram for RNaseII-GS115 and RNaseII-DPY19 using a s200 10/300 column.

A marked increase in protein expression yields was observed for many proteins in the yeast harbouring *Ce*DPY19 relative to the parental *GS115* strain. The *C. elegans* UNC5 TSR and human IL21R ectodomain only expressed in the presence of *Ce*DPY19, providing excellent yields of 14 mg/L and 10 mg/L, respectively (Table 2.3): no protein was obtained from the *GS115* strain (Table 2.3). This suggests that tryptophan C-mannosylation is required for the folding, trafficking and/or stability of these proteins. Similarly, a four-fold increase in protein production was observed for the human EPOR ectodomain and the mindin TSR in the presence of *Ce*DPY19 (Table 2.3). For IL12B, there was a two-fold increase in protein secretion for protein produced in the *Ce*DPY19 strain (4 mg/L) as compared to protein produced in the *GS115* strain (2 mg/L) (Table 2.3). RNaseII also showed a two-fold increase in secretion for protein produced in the presence of *Ce*DPY19 (Table 2.3). There was no difference in protein production observed for the human PLR ectodomain (Table 2.3). Unfortunately, EBOV sGP did not express in neither the *GS115* or in the presence of *Ce*DPY19 (Table 2.3). Perhaps this the protein construct can be optimised for the future production of this protein. Proteins produced in the DPY19 and *GS115* strains appeared identical by SDS-PAGE (Figure 2.6A),

consistent with the fact that tryptophan mannosylation adds only 162 Da to the protein and this is just a small fraction of the protein's molecular mass.

Table 2. 3: Summary of proteins produced in the *GS115* and DPY19 strains.

Protein	Species	Isolated yield (mg/ml) GS115	Isolated yield (mg/ml) CeDPY19	Enzyme	Glycopeptides	%
UNC5	*C. elegans*	n/d	**14**	Trypsin	LDGG**W**SS**W**SD**W**SACSSSCHR LDGG**W**SS**W**SD**W**SACSSSCHR LDGG**W**SS**W**SD**W**SACSSSCHR LDGG**W**SS**W**SD**W**SACSSSCHR	n/d <0.1 **51** **49**
Mindin	*H. sapiens*	2	**7**	Trypsin	EQIGGTSDCEVSL**W**SS**W**GLCGGHCGR EQIGGTSDCEVSL**W**SS**W**GLCGGHCGR EQIGGTSDCEVSL**W**SS**W**GLCGGHCGR	<0.1 11 **89**
EPOR	*H. sapiens*	5	**20**	LysArg-N	RMAEPSFGGF**W**SA**W**SEPVSLLTPSDLDHHHHHH RMAEPSFGGF**W**SA**W**SEPVSLLTPSDLDHHHHHH	30 **70**
IL21R	*H. sapiens*	n/d	**10**	Clostripain	AGPMPGSSYQGT**W**SE**W**SDPVIFQTQSEELKEHHHHHH AGPMPGSSYQGT**W**SE**W**SDPVIFQTQSEELKEHHHHHH AGPMPGSSYQGT**W**SE**W**SDPVIFQTQSEELKEHHHHHH	<1 20 **>75**
PLR	*H. sapiens*	6	6	Lys-C	PDHGY**W**SA**W**SPATFIQIPSDFTMNDHHHHHH	100
IL12B	*H. sapiens*	3	**6**	Lys-C	AQDRYYSSS**W**SE**W**ASVPCS AQDRYYSSS**W**SE**W**ASVPCS AQDRYYSSS**W**SE**W**ASVPCS	3 **4** **93**
RNaseII	*H. sapiens*	2	**4**	Glu-C	NLYFQGKPPQFT**W**AQ**W**FE NLYFQGKPPQFT**W**AQ**W**FE NLYFQGKPPQFT**W**AQ**W**FE	<0.1 **>99** <0.1

2.2.5 – Analysis of recombinant protein glycosylation using mass spectrometry

The proteins produced in yeast with and without *Ce*DPY19 were analysed by mass spectrometry to assess the degree to which they were *C*-mannosylated. Each protein was proteolytically digested with an appropriate enzyme (Table 2.3) to generate a peptide spanning the WXXW consensus motif that could be analysed by LC-MS/MS. LC-MS and semi-quantitation using extracted ion-counts (EIC) provided a means to approximate the relative abundance of each glycoform in the protein samples. Collision-induced dissociation (CID) and higher-energy collision dissociation (HCD) MS/MS methods were then used to sequence each peptide and localise the *C*-mannosylation sites.

IL21R produced in the presence of *Ce*DPY19 was digested with clostripain to provide the peptide ^{202}AGPMPGSSYQGTWSEWSDPVIFQTQSEELKE^{232}HHHHHH, which was observed in the unmodified, mono-hexosylated and di-hexosylated forms (Figure 2.7A). The mono-hexosylated peptide at RT = 30.53 min accounted for ≈23% of total peptide and HCD-MS/MS localised the *C*-mannosylation site to W^{214}, the first W of the WXXW motif (Figure

2.7B). The di-hexosylated peptide at RT = 29.42 min accounted for ≈76% of total peptide and HCD-MS/MS localised the *C*-mannosylation sites to W^{214} and W^{217} (Figure 2.7C). The unmodified peptide at RT = 31.65 min accounted for <1% of total peptide. No IL21R was obtained in the absence of *Ce*DPY19 and so comparable analysis of protein lacking *C*-mannosylation was not possible.

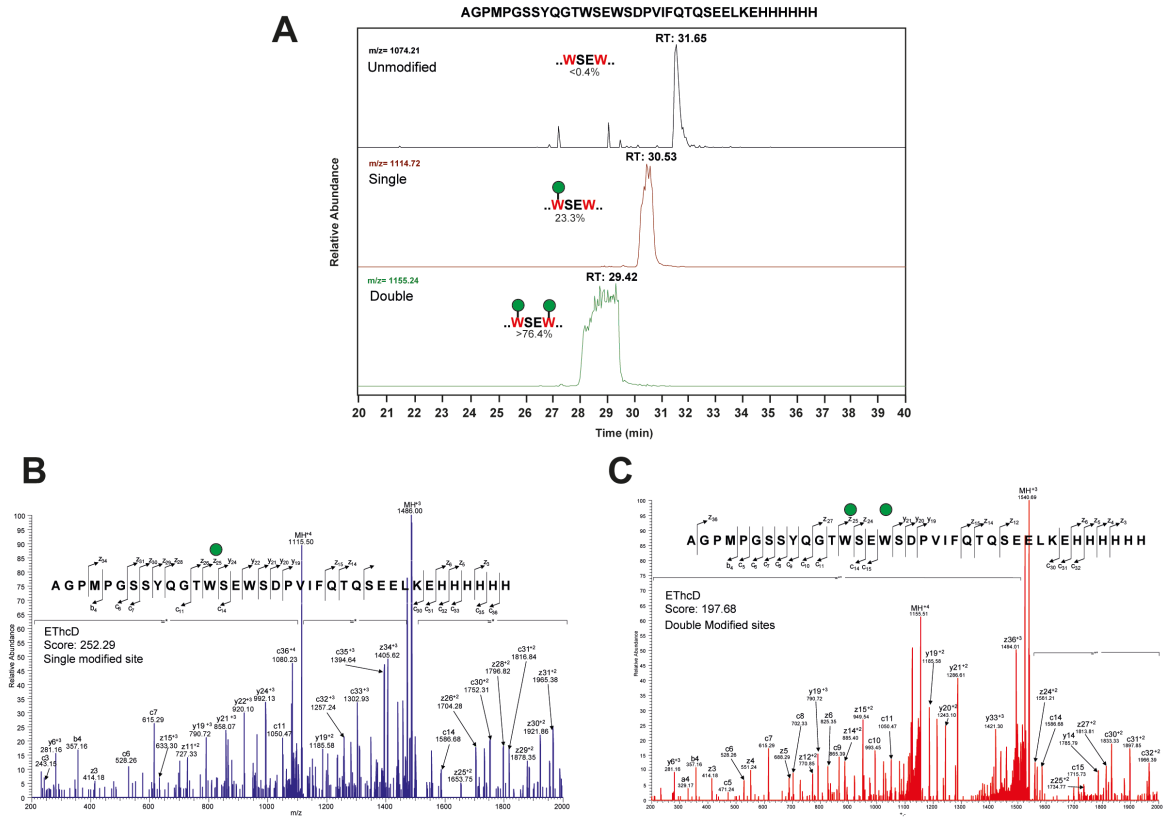


Figure 2. 7: Interleukin 21 receptor (IL21R) is *C*-mannosylated on both Trp214 and Trp217.

a) Extracted ion chromatograms of the observed glycoforms of the peptide of interest for IL21R produced in the DPY19 strain. Il21R was digested with clostripain. Green circle represents a mannose molecule.
b) HCD MS/MS spectrum for singly *C*-mannosylated IL21R peptide
c) HCD MS/MS spectrum for doubly *C*-mannosylated IL21R peptide.

EPOR produced in the presence of *Ce*DPY19 was reduced, alkylated and digested with LysArgiNase to produce a peptide ^{223}RMAEPSFGGFWSAWSEPVSLLTPSDLD^{249} which was only observed in the unmodified and singly *C*-mannosylated form (Figure 2.8A, B). EPOR was *C*-mannosylated on Trp^{233} but is not modified on the second tryptophan residue and the

singly modified form accounts for ≈70% of total peptide while the unmodified form accounts for ≈30% of total peptide (Table 2.3). Similar analysis of EPOR produced in the absence of *Ce*DPY19 showed no evidence of tryptophan mannosylation. Given that there was no change in secretion levels between PRLR produced in the GS115 strain and the presence of *Ce*DPY19 we opted to examine both of these samples using intact LC-MS. Figure 2.8C and 2.8D show that there is no major difference between the molecular weights of PRLR produced in the *GS115* strain (27695 Da) and in the presence of *Ce*DPY19 (27678 Da). We then decided to investigate the peptide fragmentation profile of PRLR produced in both strains using Lys-C and figure 2.8C and 2.8D show that both the samples produced identical fragmentation pattern providing further evidence that PRLR produced in the presence of *Ce*DPY19 is not *C*-mannosylated.

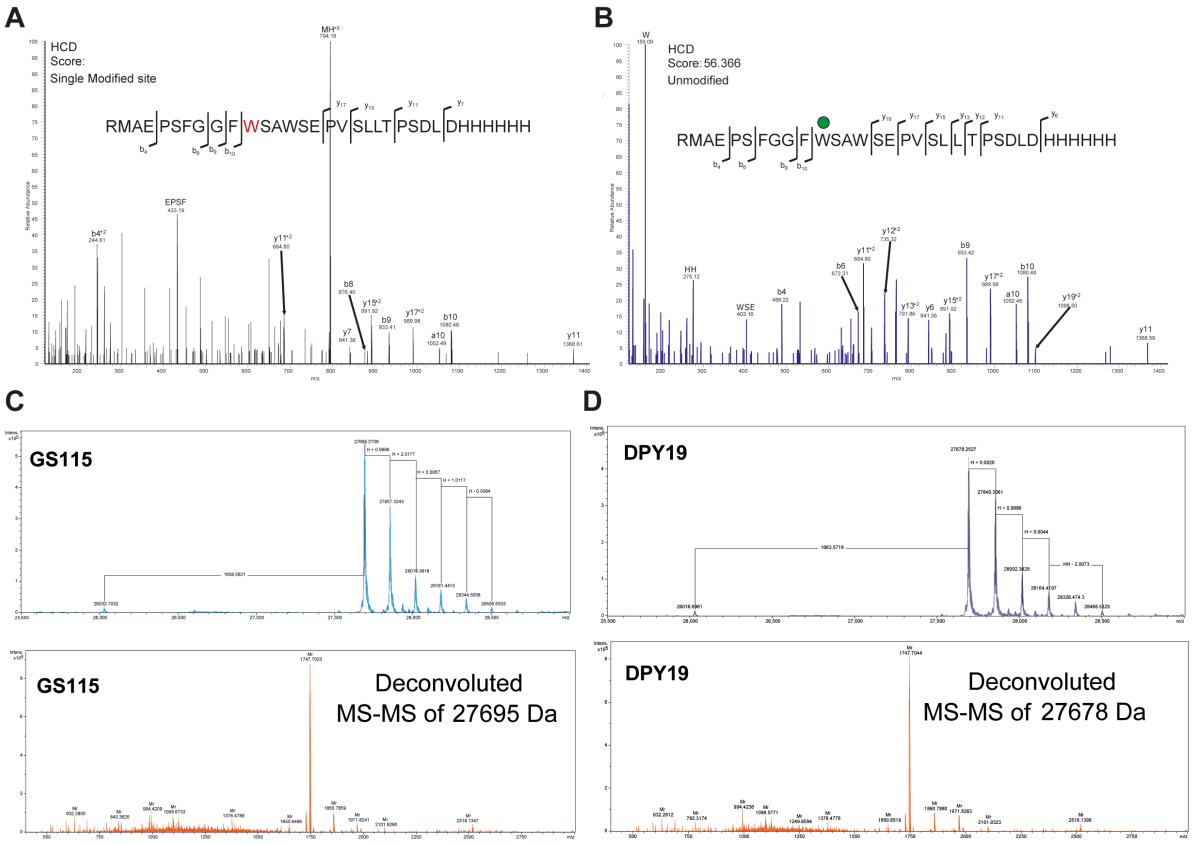


Figure 2. 8: Erythropoietin receptor (EPOR) is singly modified on Trp233 and prolactin receptor (PRLR) is not *C*-mannosylated.

a) HCD MS/MS spectrum for unmodified EPOR

b) HCD MS/MS spectrum for singly modified EPOR

c) Extracted ion chromatogram and (CID) MS/MS of PRLR produced in the *GS115* strain

d) Extracted ion chromatogram and (CID) MS/MS of PRLR produced in the DPY19 strain

IL12B produced in the presence of *Ce*DPY19 was reduced, alkylated, and treated with Lys-C protease to provide the peptide of interest ^{309}AQDRYYSSSWSEWASVPCS^{328}, which was observed by LC-MS in the unmodified, mono-hexosylated and di-hexosylated forms (Figure 2.9A). The unmodified peptide was observed at RT = 49.15 min and accounted for ≈3% of total peptide (Figure 2.9B). The mono-hexosylated peptide at RT = 43.96 min was only ≈ 4% abundance with *C*-mannosylation at W^{318} (Figure 2.9C). The majority of the peptide was in the di-hexosylated form at RT = 36.98 min with an abundance of ≈93% and *C*-mannosylation at W^{318} and W^{321} (Figure 2.9D). Similar analysis of IL12B produced in the absence of *Ce*DPY19 showed no evidence of tryptophan mannosylation.

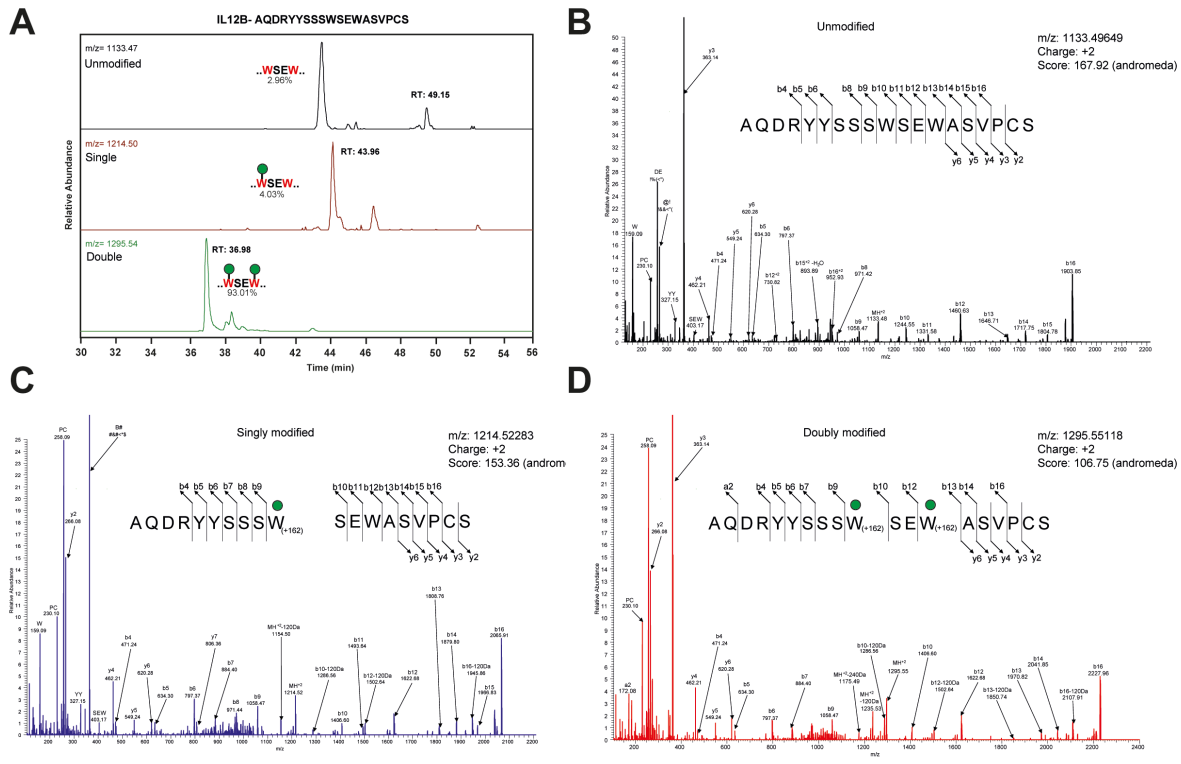


Figure 2. 9: Interleukin 12B (IL12B) is *C*-mannosylated predominantly on Trp318 and Trp321.

a) Extracted ion chromatograms of the observed glycoforms of the peptide of interest for IL12B produced in the DPY19 strain. Il12B was digested with Lys-C protease. Green circle represents a mannose molecule.

b) HCD MS/MS spectrum for unmodified IL12B peptide

c) HCD MS/MS spectrum for singly *C*-mannosylated IL12B peptide.

d) HCD MS/MS spectrum for doubly *C*-mannosylated IL12B peptide.

UNC5 produced in the presence of *Ce*DPY19 was reduced, alkylated, and treated with trypsin to provide the peptide ^{83}LDGGWSSWSDWSACSSSCHR^{292}. LC-MS could scarcely detect the unmodified peptide at RT = 46.93 min and the singly modified peptide at RT = 41.49 min, with both accounting for <0.1% of the total peptide (Figure 2.10A). HCD-MS/MS was used to fragment the singly modified peptide to localise the site of glycosylation to W^{290} (Figure 2.10B). The doubly modified peptide at RT = 36.30 min comprised ≈54% (Figure 2.10A) of total peptide and glycosylated on W^{287} and W^{290} (Figure 2.10C). The triply modified peptide at RT = 30.01 min accounted for ≈45% of total peptide (Figure 2.10A) and had *C*-mannosylation at W^{287}, W^{290} and W^{293} (Figure 2.10D). No UNC5 was obtained in the absence of *Ce*DPY19 and so comparable analysis of protein lacking *C*-mannosylation was not possible.

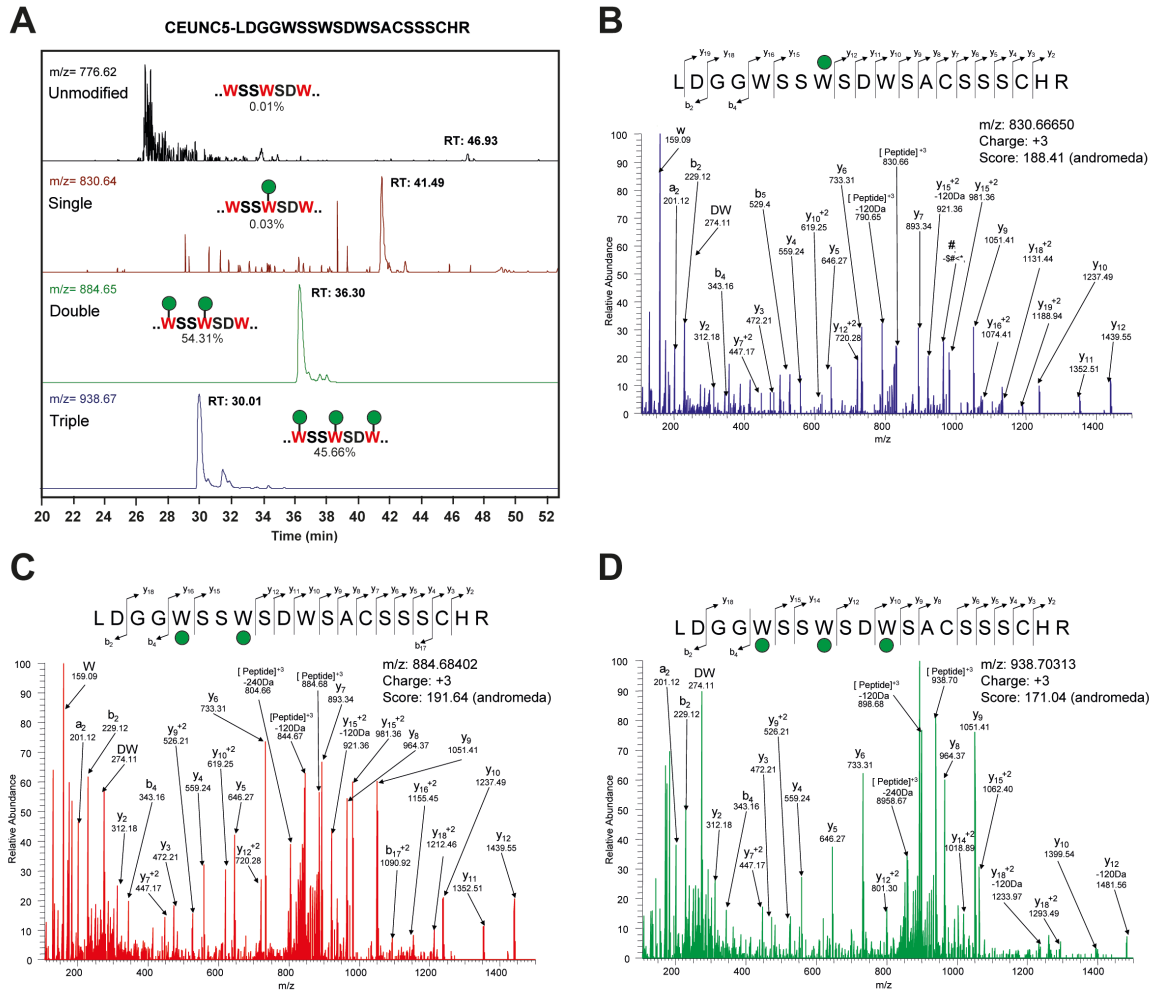


Figure 2. 10: UNC5 is *C*-mannosylated on Trp287, Trp290 and Trp293.

a) Extracted ion chromatograms of the observed glycoforms of the peptide of interest for UNC5 produced in the DPY19 strain. UNC5 was digested with trypsin. Green circle represents a mannose molecule.
b) HCD MS/MS spectrum for singly *C*-mannosylated UNC5 peptide.
c) HCD MS/MS spectrum for doubly *C*-mannosylated UNC5 peptide.
d) HCD MS/MS spectrum for triply *C*-mannosylated UNC5 peptide.

Mindin produced in the presence of *Ce*DPY19 was reduced, alkylated and proteolysed using trypsin to provide the peptide EQIGGTS^{277}DCEVSLWSSWGLCGGHCGR^{295}, which LC-MS/MS revealed to be present in the unmodified, mono- and di- hexosylated forms (Figure 2.11A). The mono-hexosylated peptide at RT = 48.10 min accounted for ≈15% of total peptide count and *C*-mannosylation was localised to W^{283} (Figure 2.11B). The di-hexosylated tryptic peptide at RT = 41.79 min accounted for ≈84% of total peptide and *C*-mannosylation sites were localised to W^{283} and W^{286} (Figure 2.11C). The unmodified form of the peptide at RT = 52.54 min accounted for <0.1% of total peptide. Similar analysis of mindin produced in the absence of *Ce*DPY19 showed no evidence of tryptophan mannosylation.

RNaseII produced in the presence of *Ce*DPY19 was treated with Glu-C protease to provide the peptide NLYFQG^{28}KPPQFTWAQWFE^{39}, which LC-MS analysis revealed to be present as the unmodified, mono-hexosylated and di-hexosylated forms (Figure 2.12A). The majority of the peptide was observed in the mono-hexosylated form at RT = 68.49 min and accounted for ≈99.8 % of the sample and was *C*-mannosylated on W^{34} (Figure 2.12B). The di-hexosylated form at RT = 62.23 min accounted for only ≈0.1% and *C*-mannosylation occurred at W^{34} and W^{37}. The unmodified form of the peptide at RT = 74.42 min only accounted for ≈0.1% of the sample.

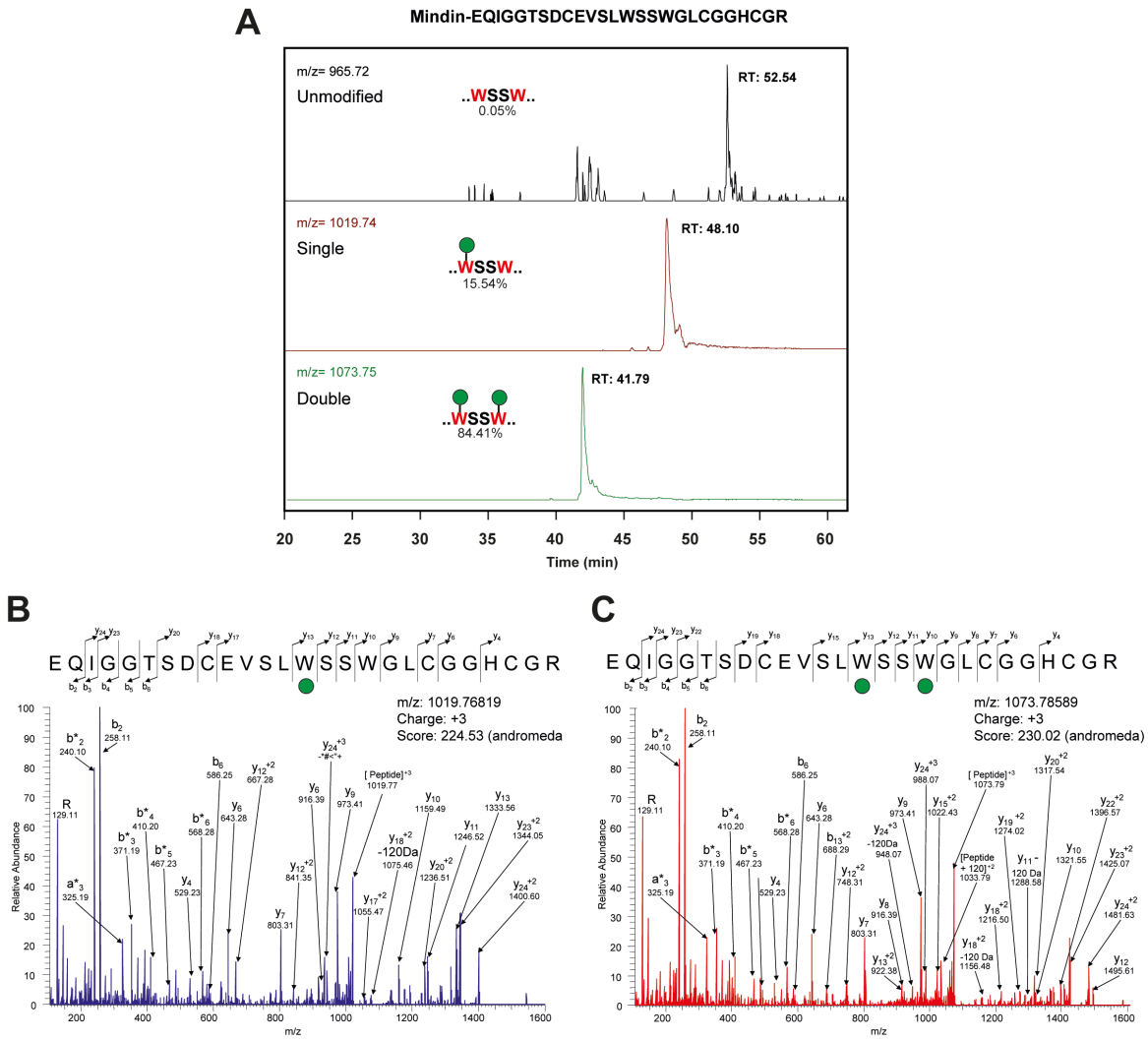


Figure 2. 11: Mindin is *C*-mannosylated on Trp283 and Trp286.

a) Extracted ion chromatograms of the observed glycoforms of the peptide of interest for mindin produced in the DPY19 strain. Mindin was treated with trypsin. Green circle represents a mannose molecule.

b) HCD MS/MS spectrum for singly *C*-mannosylated mindin peptide.

c) HCD MS/MS spectrum for doubly *C*-mannosylated mindin peptide.

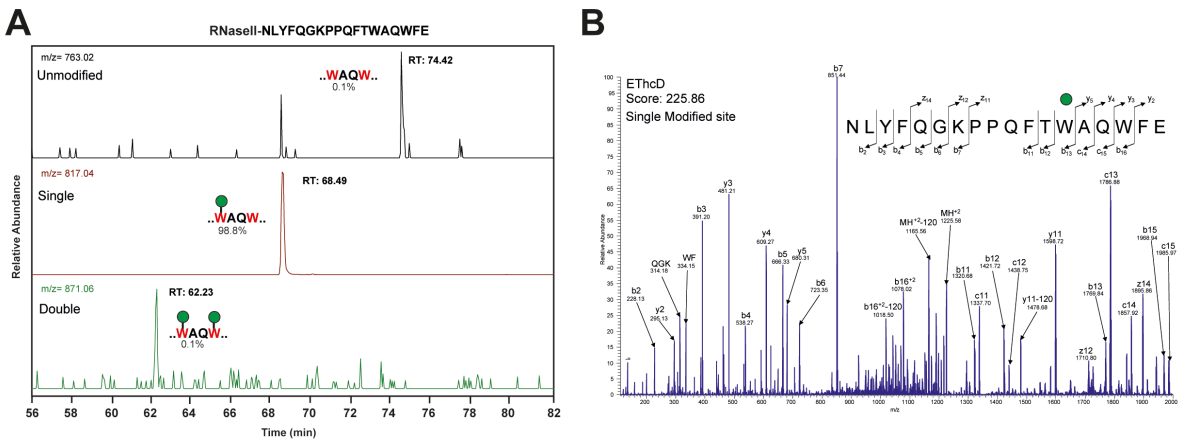


Figure 2. 12: RNaseII is predominantly *C*-mannosylated on Trp34.

a) Extracted ion chromatograms of the observed glycoforms of the peptide of interest for RNaseII produced in the DPY19 strain. RNaseII was digested with Glu-C protease. Green circle represents a mannose molecule.

b) HCD MS/MS spectrum for singly *C*-mannosylated RNaseII peptide

2.3 - Discussion

There are over 500 proteins in the human proteome that contain the WXXW motif and pass the secretory pathway, making them potential substrates for tryptophan mannosylation [13]. Currently, just over 20 human proteins have been found with *C*-mannosyl-tryptophan though there is no doubt that this number will increase with the passage of time. The impact of this modification on protein fold and function remains unknown. In this chapter, a functional *C*-mannosylation pathway has been engineered into the yeast *P. pastoris* to enable the study of DPY19 glycosyltransferases and the production of recombinant proteins with and without tryptophan mannosylation to probe the biological role(s) of this enigmatic protein modification.

This work compliments on-going efforts to 'humanise' the glycosylation pathways of *P. pastoris* to provide a cheap and efficient protein production platform for the biopharmaceutical industry. Producing proteins with human-like glycosylation is critical to ensure the proper function of the protein, optimise stability and pharmacokinetics and minimise antigenicity [150]. Although yeast ordinarily hyper-mannosylate proteins, they can accommodate dramatic changes to their glycosylation pathways and are easy to genetically manipulate. *P. pastoris*

strains with humanised *N*-glycosylation already exist and have been used in the biotechnology industry [134, 136]. *O*-linked glycosylation has also been engineered in *P. pastoris* to produce proteins with human-like sialylated *O*-linked glycans [141]. The present work, while more modest in scope than these efforts, compliments *N*- and *O*-glycan engineering to provide human-like *C*-mannosylation in *P. pastoris*.

How well the glycosylation patterns on the human proteins we have produced here emulate those of endogenous human proteins is difficult to assess because so few endogenous proteins have had their *C*-mannosylated examined. Indeed, RNaseII is the only protein isolated from human samples (urine) to be analysed for *C*-mannosylation: all other samples examined in the literature are recombinant proteins produced in human cell lines (e.g. HEK293). RNaseII purified from human urine is exclusively singly *C*-mannosylated. Our recombinant RNaseII expressed in the presence of *Ce*DPY19 is 99.8% singly *C*-mannosylated [8]. While this is only a single data point, it is suggestive that this engineered system produces proteins with human-like patterns of tryptophan *C*-mannosylation.

Our protein expression data suggest that *C*-mannosylation aids in the proper folding and/or trafficking of some, but not all, proteins with a WXXW consensus motif. Previous efforts to probe the function of *C*-mannosylation involved mutating tryptophan residues in the WXXW motif to phenylalanine and alanine [21, 45, 57, 71]. The WXXW motif in many of these proteins forms π-cation tryptophan ladder motif, which involves the planar stacking arrangement of the guanidinium moiety of arginine over the centre of the aromatic ring of tryptophan to stabilise protein fold [19]. As such, it is difficult to interpret what these mutation studies say about tryptophan mannosylation. Nonetheless, such mutations in the WXXW motif of IL21R prevented the receptor from being trafficked to the cell surface [21], which mimics what we observe for IL21R in yeast lacking *Ce*DPY19. Collectively this data implies that tryptophan *C*-mannosylation is important for the folding and/or trafficking of IL21R. Similar studies have been done for the thrombopoietin receptor (TPOR), common gamma chain, and Punctin-1: mutations in the tryptophan residues of the WXXW motif resulted in impaired trafficking to the cell surface [57] [63]. It would be interesting to see if our system could verify these results by examining the impact of mannosylation in the absences of mutations to protein sequence.

Interestingly, not all proteins with a WXXW are modified by *C*-mannosyltransferases, even when both are expressed at high levels, as was the case for the prolactin receptor (PRLR) in our yeast bearing *Ce*DPY19. Both the parental *GS115* and engineered yeast lines produced the same amount of unglycosylated protein, revealing that the WXXW consensus motif is necessary but not sufficient for tryptophan mannosylation. We rationalise this observation by postulating that the DPY19 enzyme only acts on unfolded or misfolded polypeptides in a co-translational sense such that proteins that rapidly attain their native fold without the need for glycosylation will not be available for modification by the DPY19 enzyme.

This simple examination of differences in protein expression levels helps to build the narrative that tryptophan *C*-mannosylation play a significant role in protein folding. The means to produce protein with and without the modification enables a more quantitative study of impacts on protein stability and function, which is addressed in the next chapter.

The mass spectrometry data compiled here also begins to provide insights into the substrate preferences of the *Ce*DPY19 enzyme. It has been established that the tryptophan mannosyltransferases preferentially modify the N-terminal Trp of the WXXW motif. Close examination of the data for UNC5 reveals that, of the three possible mannosylation sites in the WSSWSDW peptide, it is the second site that is preferentially modified, followed by the first site. Could it be then that the WSDW motif binds preferentially to *Ce*DPY19 over the WSSW peptide? The substrate preferences and tolerances of the *Ce*DPY19 enzyme is something that can be conveniently studied with the yeast system and will be addressed in later chapters.

2.4 - Conclusions

This chapter has described an engineered *P. pastoris* that enables the production of recombinant proteins with tryptophan *C*-mannosylation. This could be useful for the production of biologics like IL12B and vaccines against malaria or, with more optimisation, the Ebola virus. In at least one case, recombinant protein produced in this system had identical *C*-mannosylation to endogenous human samples. Protein expression data from this yeast system supported the idea that tryptophan mannosylation plays a role in protein folding and trafficking. Having the means to produce recombinant proteins with and without *C*-

mannosylation now provides the means to examine this in great detail, explore the impact *C*-mannosylation has on protein function and generate of antibody tools to detect the modification. This system also enables further study of the DPY19 enzyme by enabling the facile introduction of mutations to the enzyme or its substrates. All of the ideas are probed in the following chapters.

Chapter 3

Insights into the function of tryptophan *C*-mannosylation

3.1 - Introduction

Since the discovery of the first C^2-α-D-mannosyl-tryptophan on RNaseII in 1994, the function of this unusual modification has remained elusive [8]. The reason for this is because the genetic identity of the enzyme that catalyses this reaction remained a mystery. This was until the discovery of the first *C*-mannosyltransferase gene, *dpy19*, in *C. elegans* in 2013 which paved the way for study of this modification [68]. Previous studies have sought to elucidate a role for tryptophan *C*-mannosylation through mutagenesis of the tryptophan residues in the WXXW consensus motif to phenylalanine and alanine [21, 57, 63, 71]. Mutagenesis of the *C*-mannosylated WSEW motif in IL21R to FSEW perturbed protein trafficking, resulting in accumulation of the receptor in the ER and very little IL21R at the cell surface [21]. Similar results have been observed for IL2RG, TPOR, R-spondin and Punctin [21, 57, 63, 71]. It remains unclear as to whether or not these trafficking defects are a result of the amino acid substitutions, a loss of *C*-mannosylation, or both.

In all of these proteins, the WXXW motif is involved in the formation of a tryptophan ladder motif, which is found in both the extracellular domains of type I cytokine receptors and TSR domains [17, 18]. Thus, it is likely that these Trp to Phe or Ala substitutions play a significant role in destabilising these proteins. This would mean that the lack of secretion of these proteins in these previous studies were not a result of a lack of *C*-mannosylation. Indeed, similar mutations on *C*-mannosylated proteins lacking a tryptophan ladder, such as lipoprotein lipase (LPL), actually resulted in an apparent reduction of LPL activity [97]. Though it should be noted that this data was collected on crude cell lysate without the determination of Michaelis-Menten parameters, making it difficult to draw meaningful conclusions about the role of *C*-mannosylation on LPL function.

In Chapter 2 of this thesis, I took advantage of the discovery of the *dpy19* gene and integrated it into *P. pastoris*. This enabled the production of proteins with and without the modification,

which provides an opportunity to examine the function of this modification on a number of proteins without any confounding mutagenesis. One of these proteins was the TSR domain of mindin.

Mindin (Spondin-2) is an extracellular matrix protein belonging to the F-spondin protein family and with a variety of different functions (Figure 3.1A) [75-80]. Mindin contains a TSR domain and a spondin domain, where the spondin domain is known to bind to $\alpha_M\beta_2$ integrin on macrophage cell surfaces [69]. The primary function of mindin is to serve as a pathogen recognition molecule where it recognises LPS on bacterial cell surfaces and serves as an opsin for phagocytosis by macrophages [81]. There have been studies conducted to elucidate which domain of mindin binds to LPS and lipid A, but this has not been successful because the TSR domain could not be expressed.

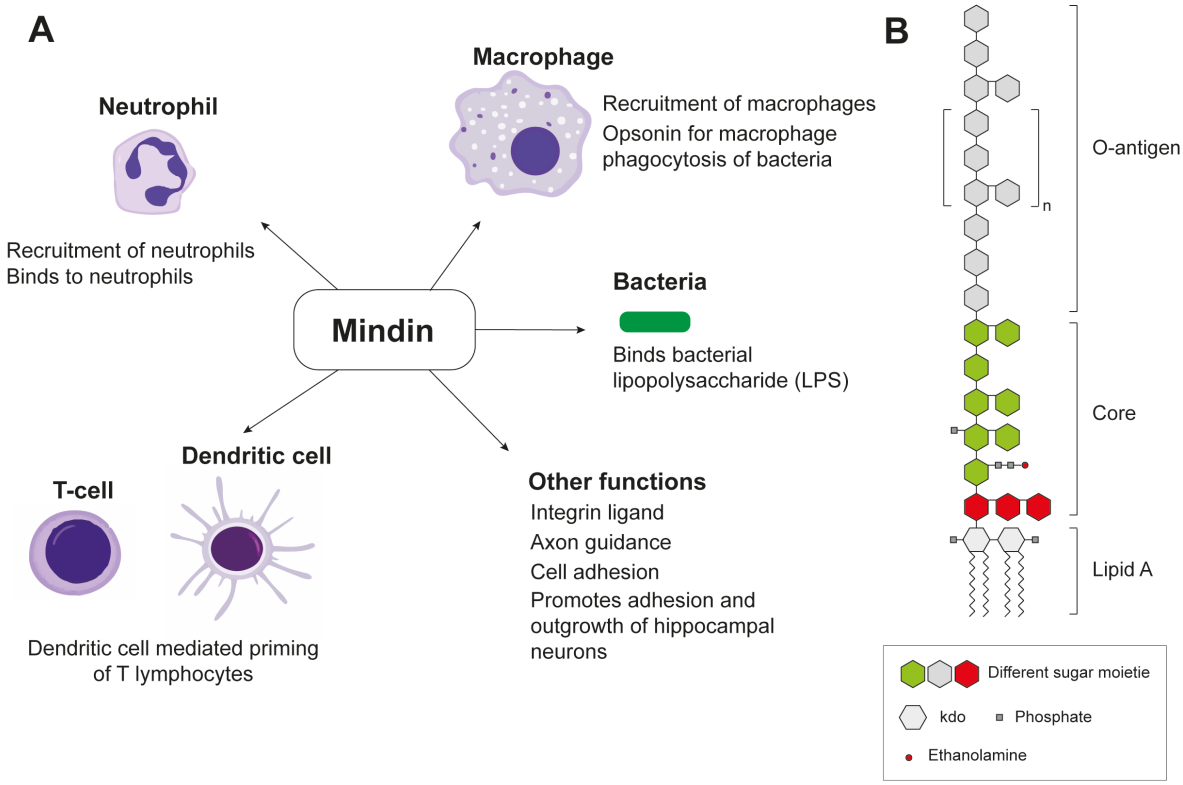


Figure 3. 1: Functions of Spondin-2 (mindin) in humans through its effects on different cells and organisms

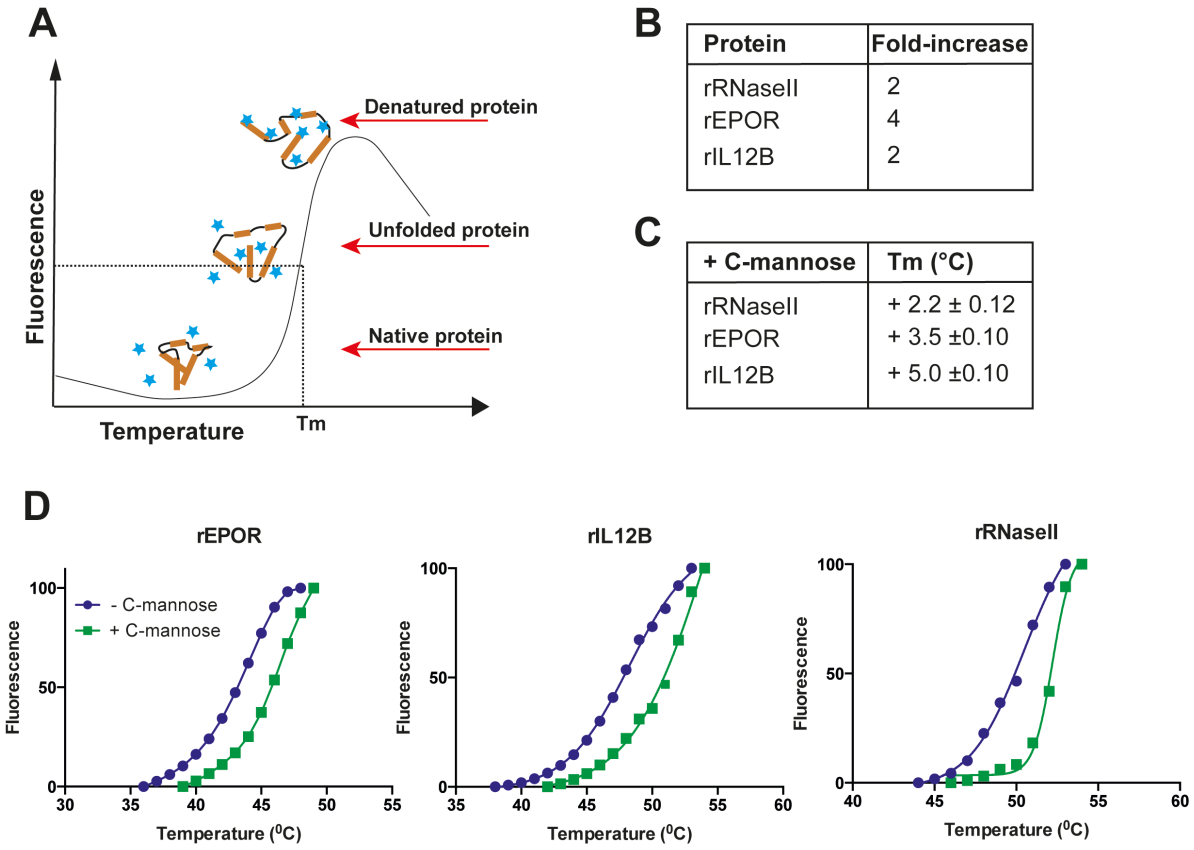
a) Diagram illustrating the myriad of roles that mindin plays for certain cell types.
b) General structure of bacterial LPS (*E. coli*) showing the O-antigen, core and Lipid A regions.

This Chapter examines the role of *C*-mannosylation on the stability and function of mindin and other proteins.

3.2 - Results

3.2.1 - *C*-mannosylation stabilises proteins to thermal denaturation

In Chapter 2, we expressed a number of proteins in *Pichia pastoris* in the presence and absence of DPY19 to produce proteins with and without *C*-mannosylation, respectively. The yields of many proteins were higher when *C*-mannosylated: on average protein yields were 2-fold higher in the presence of DPY19. Interestingly, UNC5 and IL21R did not secrete at all without *C*-mannosylation, suggesting that these proteins failed to fold or were unstable without the C^2-α-D-mannosyl-tryptophan. Differential scanning fluorimetry (DSF) provides a convenient means to examine the thermal stability of proteins [151]. DSF works by gradually heating proteins in the presence of a dye such as SYPRO orange, which fluoresces more strongly when it binds to hydrophobic amino acids that become solvent exposed upon denaturation (Figure 3.2A). RNaseII, IL12B and EPOR with and without C^2-α-D-mannosyl-tryptophan were analysed by DSF. RNaseII, EPOR and IL12B with C^2-α-D-mannosyl-tryptophan had a 2, 4, and 2- fold increase in their secretion, respectively, as compared to protein without C^2-α-D-mannosyl-tryptophan (Figure 3.2B). Tryptophan mannosylation increased the thermal stability of RNaseII by 2.2 ± 0.12 °C (Figure 3.2C, D), EPOR by 3.5 ± 0.1 °C, and IL12B by 5.0 ± 0.1 °C (Figure 3.2C, D).



Protein	Fold-increase
rRNaseII	2
rEPOR	4
rIL12B	2

+ C-mannose	Tm (°C)
rRNaseII	+ 2.2 ± 0.12
rEPOR	+ 3.5 ±0.10
rIL12B	+ 5.0 ±0.10

Figure 3. 2: *C*-mannosylation stabilises proteins and is required for the proper folding and secretion of target proteins.

a) Diagram depicting the fold-state of protein undergoing thermal denaturation and binding of dye (blue stars) to hydrophobic residues buried in the protein structure.

b) Table showing a summary of fold-increase in protein secretion for rRNaseII, rEPOR, rIL12B for protein produced with C^2-α-D-mannosyl-tryptophan.

c) Table showing a summary of the difference in melting temperature (T_m) of proteins with a C^2-α-D-mannosyl-tryptophan as compare to protein without. Data is presented as mean ± standard deviation with triplicates.

d) Thermal denaturation profiles for rRNaseII, rEPOR, rIL12B with and without C^2-α-D-mannosyl-tryptophan. The denaturation profiles are representations of the average of three replicates. 2 µg of rRNaseII, rEPOR, rIL12B was used in the assay. The data was trimmed to give a sigmoid and it was then normalised and a Boltzmann sigmoidal curve was applied to the data.

3.2.2 - *C*-Mannosylation does not impact RNaseII activity

RNaseII is a protein with myriad functions, though its most important property is its RNase activity. Other functions include its ability to activate dendritic cells and serve as a chemoattractant for these cells, its anti-viral activity and its ability to activate toll-like receptor 2 (TLR2) [152-156]. Tryptophan mannosylation may play a role in a number of these functions, though the simplest to assay for is the RNase activity. The RNase assay utilised here takes advantage of ethidium bromide's fluorescence at 450 nm when bound to RNA [157]. I first titrated the RNA while keeping ethidium bromide concentration constant to find a RNA concentration range that gave a linear response (Figure 3.3A). When RNaseII hydrolyses these RNA molecules it disrupts binding of the ethidium bromide, resulting in a decay in florescence that can be used to determine enzyme kinetics (Figure 3.3B). Working within this RNA concentration range, I examined the kinetic activity of RNaseII with and without *C*-mannosylation by keeping the enzyme concentration and ethidium bromide constant while titrating RNA concentrations from 0 to 60 μg/ml. In this concentration range, we do not see saturation of the enzyme with substrate in the Michaelis-Menten curves, which precludes us from determining k_{cat} and K_M values individually, only k_{cat}/K_M values. RNaseII-GS115, which lacked C^2-α-D-mannosyl-tryptophan had a k_{cat}/K_M value of 2.3 ± 0.2 ml μg^{-1} s^{-1} (Figure 3.3C). RNaseII-DPY19, which possessed a single C^2-α-D-mannosyl-tryptophan, had a k_{cat}/K_M value of 2.0 ± 0.3 ml μg^{-1} s^{-1} (Figure 3.3D). These values are within standard error of each other, suggesting that *C*-mannosylation is not important for the catalytic activity of RNaseII.

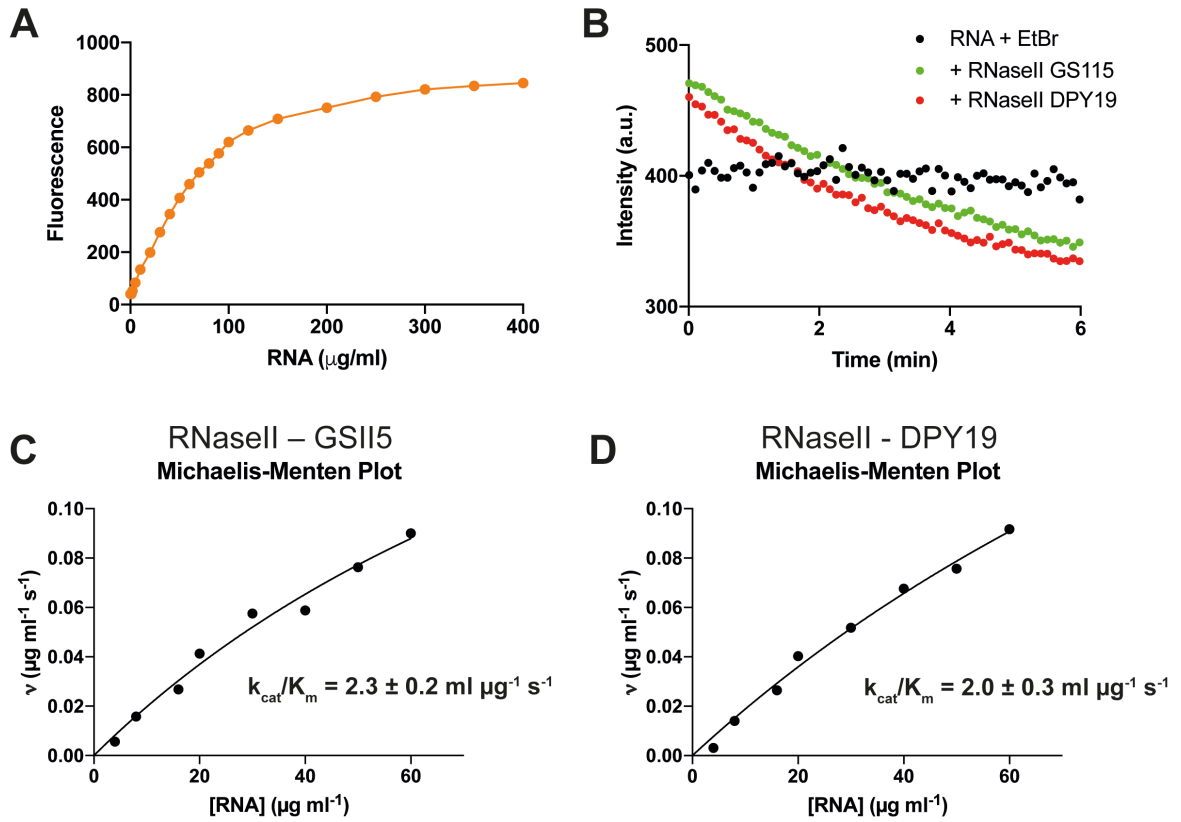


Figure 3. 3: *C*-mannosylation stabilises secreted RNaseII but is not required for catalytic activity

a) Dose-response curve for ethidium bromide (4 µM) with different concentrations of RNA (0-400 µg/ml).

b) Decay of fluorescence once RNaseII-GS115 (green) and RNaseII-DPY19 (red) is added to ethidium bromide and RNA (black) over time.

c) Michaelis-Menten plot for RNaseII-GS115

d) Michaelis-Menten plot for RNaseII-DPY19

3.2.3 - The TSR domain of Spondin-2 (mindin) mediates binding to LPS and this activity is influenced by tryptophan mannosylation

Mindin (Spondin-2) is an extracellular matrix protein with several different functions [75-80]. Its primary function is to serve as a pathogen recognition molecule, where it recognises LPS on bacterial cell surfaces and serves as an opsin for phagocytosis by macrophages [81]. Mice lacking mindin cannot clear pathogens and macrophages deficient in mindin show defective responses to microbial stimuli [81]. Mindin is comprised of two domains, a spondin domain (31-221) and a TSR domain (277-331) in which W^{283} and W^{286} of the TSR domain are *C*-

mannosylated (Figure 3.4A). Full-length mindin produced in mammalian cells shows significant binding to LPS, while just the spondin domain from mammalian cells or the full-length mindin produced in bacterial cells has very little binding to LPS. Together, this suggests that the TSR domain of mindin and tryptophan mannosylation may be important for the LPS-binding properties of mindin.

The means to produce the TSR domain of mindin with and without C^2-α-D-mannosyl-tryptophan provided an opportunity to explore these questions. Attempts to examine the thermal stability of mindin with and without tryptophan mannosylation were thwarted by the inability of SYPRO orange to provide a meaningful thermal denaturation profile (data not shown). Some proteins are not amendable to DSF because they contain exposed hydrophobic residues that bind to SYPRO orange in both the folded and unfolded states, providing a saturated signal even before denaturation.

To test the hypothesis that the TSR domain was binding to LPS and that *C*-mannosylation might play a role in this binding, we established an enzyme-linked immunosorbent assay (ELISA) for LPS binding. In this ELISA, 96-well plates were coated with the recombinant mindin TSR with and without C^2-α-D-mannosyl-tryptophan, the production of which is outlined in Chapter 2, then blocked the plates with BSA, followed by incubation with LPS (Figure 3.4B). After washing, these wells were then probed with a mouse anti-LPS antibody followed by an anti-mouse IgG antibody conjugated to HRP and developed using a standard ELISA chromogenic reagent.

Figure 3.4C shows that mindin TSR domains, with or without C^2-α-D-mannosyl-tryptophan, are necessary and sufficient for LPS binding. We examined the use of BSA as a control for binding to LPS but we saw significant binding to LPS (Figure 3.4C). BSA is known to bind to LPS when BSA is coated on plates using high pH [158]. We therefore sought to optimise the negative control and explored the use of casein which does not bind LPS. To see if *C*-mannosylation plays a role in the binding of mindin to LPS, the concentration of both mindin TSR samples were kept constant and the concentration of LPS was varied. Figure 3.4D shows that the mindin TSR without C^2-α-D-mannosyl-tryptophan shows greater binding capacity to LPS than the sample with C^2-α-D-mannosyl-tryptophan. Thus, it appears that C^2-α-D-mannosyl-tryptophan can reduce the affinity of LPS for the mindin TSR and thus regulate the activity of this protein. Questions as to which mannosylation site is most responsible for this

reduced binding and whether or not this glycosylation directly impedes LPS binding or acts through an allosteric mechanism remains to be explored.

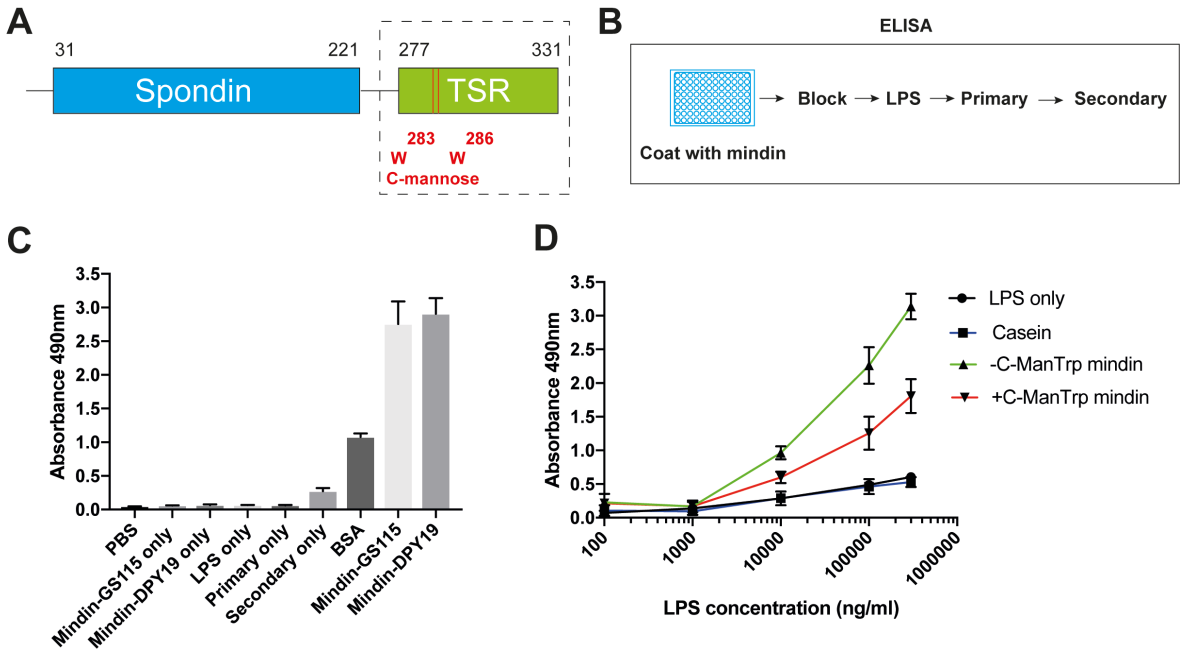


Figure 3. 4: The TSR domain of Spondin-2 (mindin) mediates binding to LPS and is required for phagocytosis and T-cell priming

a) Schematic of the domain architecture of mindin. Red lines in the TSR domain of mindin denote the sites of *C*-mannosylation (W^{283} and W^{286}).

b) Flowchart showing how the mindin-LPS enzyme-linked immunosorbent assay (ELISA) was conducted.

c) Mindin-LPS ELISA showing all appropriate controls. All data is in triplicates.

d) Dose-response curve for mindin-LPS binding. X-axis is in $\log_{10}$ scale.

3.3 - Discussion

For the first time, we show that *C*-mannosylation directly stabilises proteins by engineering a system that allows the direct comparison of proteins with and without a C^2-α-D-mannosyl-tryptophan. Our results are not confounded by disruption of the tryptophan ladder through mutagenesis of the tryptophan residues and any effect we see is a direct consequence of the presence or absence of a C^2-α-D-mannosyl-tryptophan. In chapter 2, we discovered that

proteins produced in the DPY19 strain with a C^2-α-D-mannosyl-tryptophan had on average a 2 to 5-fold increase in their secretion. For Il21R and UNC5 they would only be secreted when they were produced in the DPY19 strain with a C^2-α-D-mannosyl-tryptophan. This promoted the hypothesis that *C*-mannosylation might have been playing a stabilising role thereby facilitating proper folding and trafficking to the cell surface to be secreted. Our differential scanning fluorimetry results in this chapter give unprecedented evidence that the C^2-α-D-mannosyl-tryptophan actually stabilises proteins such as RNaseII, EPOR and IL12B.

Therapeutically, being able to modulate and control whether certain proteins go to the cells surface to conduct their function may be highly useful. For instance, IL6 is a driver for certain myeloproliferative disorders and the cancer is dependent on IL6 to proliferate and survive [159]. If *C*-mannosylation is needed for the IL6 receptor (IL6Rα) to be secreted to the cell surface, then an inhibitor that inhibits *C*-mannosylation could be beneficial for killing these cancers. Il7R is another example where it drives proliferation of T-cells and causes T-cell acute lymphoblastic leukaemia (T-ALL) [160]. Il21R is another protein where inhibiting it from travelling to the cell surface might be beneficial in rheumatoid arthritis where it is implicated in the progression of pathogenesis [161]. Another great example is UNC5C, which is a protein with two TSR domains which is involved with axon guidance, apoptosis and cell migration. A mutation in UNC5C has implicated it Alzheimer's disease [162]. The mutation causes increased neuronal cell death and increase sensitivity to beta-amyloid peptide [162]. We have shown in chapter 2 that the *C. elegans* homologue of UNC5C does not secreted if it is not *C*-mannosylated.

A 19-amino acid miniature protein containing only a tryptophan ladder showed that tryptophan ladders stabilise proteins by -5.5 kcal/mol [20]. The tryptophan ladder is a structural element containing a WXXW motif that is characterised by π-cation interactions between two tryptophan residues and an arginine residue [17] [18]. Many *C*-mannosylated proteins such as type I cytokine receptors and thrombospondin type 1 repeats contain tryptophan ladders. An interesting study would be to produce this miniaturised protein with a C^2-α-D-mannosyl-tryptophan to investigate the role of *C*-mannosylation in tryptophan ladders. It could be that *C*-mannosylation helps facilitate the stabilisation of tryptophan ladders and it would be interesting to investigate how much tryptophan ladders are stabilised by the presence of a C^2-α-D-mannosyl-tryptophan.

In light of the fact that *C*-mannosylation stabilises proteins, we wanted to further characterise the function of modification. We looked to the protein RNaseII (or Eosinophil-derived neurotoxin), where one of the functions of RNaseII is to be secreted from eosinophils, neutrophils and basophils into the blood stream and acts as an anti-microbial by targeting viral RNA. For RNaseII, we show that *C*-mannosylation stabilises the protein but has no effect on the enzymatic activity of RNaseII. Given that RNaseII may need a longer half-life in the blood stream in order for it to find and degrade viral RNA, the C^2-α-D-mannosyl-tryptophan attached to the tryptophan residue is helping it achieve this longer half-life through a shielding/stabilising effect. RNaseA is very similar in amino acid sequence with RNaseII but does not have the WXXW motif. This protein is secreted from the pancreas and does not need to travel long distances to reach its targets and may be why it is not *C*-mannosylated.

We have uncovered that the thrombospondin type I repeat (TSR) domain of mindin is responsible for binding to LPS which then leads to phagocytosis of bacterial pathogens [81]. Previous work around mindin's role in binding to LPS could not shed light on which domain of mindin was causing this interaction because they could not express the TSR domain on its own. The spondin domain is known to bind to $\alpha_M\beta_2$ integrin on macrophage cell surfaces [69]. Unfortunately, there is no known structure of the TSR domain of mindin because it has not been recombinantly expressed before. A structure of the TSR domain of mindin with and without LPS bound would give insight into how this important process works on a molecular level. A possible mechanism of how this process works is that the spondin domain of mindin binds to macrophage cells surface via $\alpha_M\beta_2$ integrin, this leaves the TSR domain exposed. Once the TSR domain comes in contact with LPS on the cells surface of bacteria, it undergoes a conformational change upon binding to LPS and sends a downstream signal to macrophages that activates them for phagocytosis. Another question to ask is why there a need for the C^2-α-D-mannosyl-tryptophan on mindin if it negatively affects its ability to bind to LPS? Is it there as a control mechanism to finely tune the amount of LPS that mindin binds to? How does that benefit humans? Perhaps the C^2-α-D-mannosyl-tryptophan's role in stabilising mindin is more important than its role in LPS recognition?

TSR domains are inherently difficult domains to recombinantly express because they are disulphide rich proteins. Most attempts at expression in *E. coli* have failed or have given very low expression levels making it impossible to crystallise these structures. With over 100 TSR domain containing proteins in the human proteome, only three of them have structures of which

only one has a crystal structure. Available in the PDB is an NMR solution structure of Spondin-1 TSR domain (1SZL), a solution structure of a properdin TSR domain (1W0R), and a crystal structure of thrombospondin-1 TSR domain 2 and 3 (3R6B). There is also a NMR solution structure of Plasmodium TRAP TSR domain (2BBX), and a crystal structure of circumsporozoite protein TSR domain (3VDJ). We now have a system where we have engineered a crucial PM into that will allow the recombinant expression of many TSR domain containing proteins. For instance, brain-specific angiogenesis inhibitor 1 (BAI) is a protein that binds to phosphatidylserine on the cell surface of cells undergoing apoptosis for recognition by macrophages for phagocytosis [163]. BAI has five TSR domains and a GPS domain that anchors the protein onto the cell membrane. The TSR domains likely mediates binding to phosphatidylserine in a similar way that the TSR domain in mindin binds to LPS. Being able to express the TSR domains of BAI in our system to crystallise them with phosphatidylserine would allow us to understand how this fundamental process occurs.

3.4 - Conclusions

In conclusion, we have used our glycoengineered yeast strain to decipher the molecular and functional role of *C*-mannosylation by producing proteins with and without the modification. We have found that *C*-mannosylation plays a stabilising role in RNaseII, EPOR and IL12B. Our work found that *C*-mannosylation does not play a role in the RNase activity of RNaseII. Furthermore, we found that the TSR domain of mindin is responsible for its LPS binding properties and that *C*-mannosylation does not impact this LPS binding ability.

Chapter 4

Mutagenic study of the *C. elegans C*-mannosyltransferase DPY19

4.1 - Introduction

The DPY19 enzymes are unusual in their ability to form carbon-carbon bonds between proteins and sugars. Much remains unknown about this irreversible process, including the enzyme's mechanism and which residues are involved in catalysis. Since the discovery of first *C*-mannosyltransferase gene, *dpy19* or DumPY-19, in *C. elegans*, very little has been said about how this enzyme might work. From a chemical standpoint, the unique carbon-carbon bond formation is very interesting.

Dpy19 was initially discovered as a *C*-mannosyltransferase based on its perceived homology to bacterial OST and eukaryotic STTA/B, which are the enzymes responsible for *N*-glycosylation [68]. The authors also concluded that since both *N*-glycosylation and *C*-mannosylation use Dol-P-Man as a sugar donor substrate, that the *C*-mannosyltransferase must reside in the ER. However, it must be stressed that the DPY19 enzymes are still markedly different to the OST enzymes at a sequence level and in their predicted membrane topology [68].

In silico analysis of human DPY19L3 using a number of different transmembrane prediction software packages has found that DPY19L3 most likely has 11 transmembrane helices [164], with the N-terminus projected into the ER lumen and the C-terminus projected into the cytoplasm. Two major non-transmembrane loops at residues 65 to 155 and 359 to 415 are predicted to project into the cytoplasm [164]. The other two major non-transmembrane loops from residue 437 to 466 and 499 to 523 are predicted to project into the ER lumen. Of course, structural studies are still needed to confirm these results. However, it is important to note that DPY19L3 is very different to the *C. elegans* DPY19 and these studies done on DPY19L3 may not inform us on how DPY19 works.

Given that nothing is known about what features of the *C*-mannosyltransferases are important for catalysis, I decided to use the yeast system developed in Chapter 2 to perform some

mutagenesis studies to examine which conserved features are important for the catalytic activities of these unusual enzymes.

4.2 - Results

4.2.1 - The DPY19 family has several conserved amino acids among different phyla

Multiple sequence alignments of diverse *dpy19* homologues from organisms were used to identify residues that were conserved across the DPY19 family. For these alignments, the recently discovered human *C*-mannosyltransferases *dpy19l1* and *dpy19l3*, along with *dpy19* homologues from organisms with only one *dpy19* gene (Figure 4.1B) were used to improve the likelihood that only active *C*-mannosyltransferases were being compared. With this analysis, eight amino acid residues were identified, including: Y75, E102, E112, E118, H375, E579, R586, and Y591 (Figure 4.1B). The location of these residues is shown with respect to the predicted topology of *C. elegans* DPY19 in figure 4.1A. Seven out of eight of these residues lay in regions predicted to be in the ER lumen, and in two regions of non-transmembrane loops and only one residue is located in the cytosolic face of the ER (Figure 4.1A).

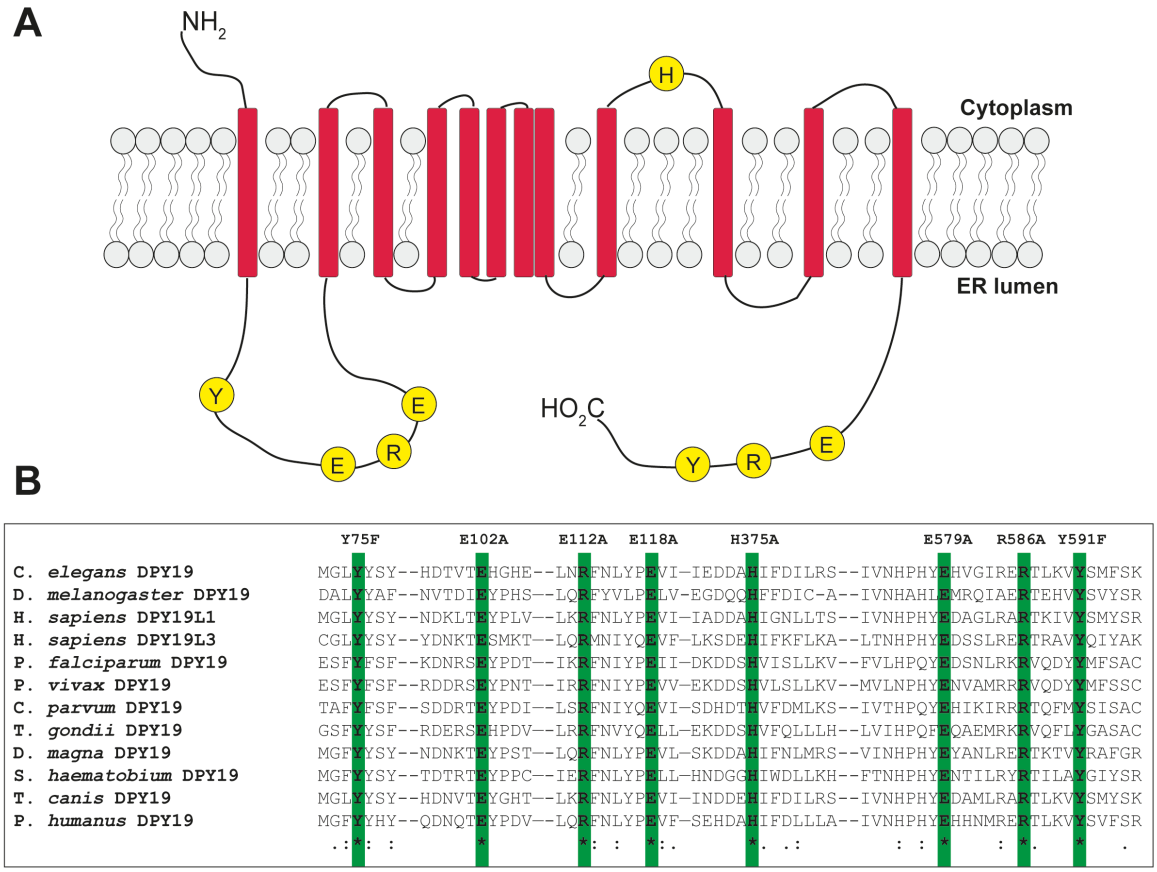


Figure 4. 1: The DPY19 gene family is highly conserved among different organisms

a) Illustration of the predicted topology of the *C. elegans* DPY19 enzyme with annotation of conserved residues.

b) Alignment and annotation of different DPY19 homologues from a wide range of organism highlighting the discovered conserved amino acid residues.

4.2.2 - A key conserved residue amongst the DPY19 family is crucial for catalysis

In order to explore the roles of these residues in catalysis, we made alanine or phenylalanine mutants of these conserved residues. Testing the activity of eight mutants using *in vitro* and cellular assays would be very time consuming. We therefore wanted to short-list mutants based on their *in vitro* activity, and take the mutants with decrease in activity and further test their activity in cells using IL12B as a reporter protein that can be examined using mass spectrometry for the extent of *C*-mannosylation (Figure 4.2A). Figure 4.2B shows that the cassette encoding these mutants have integrated into the *P. pastoris* genome. In Figure 4.2C, we can see the location of these residues and the transmembrane and non-transmembrane regions of DPY19. Figure 4.2D shows western blots of the DPY19 mutants, providing evidence that they are

expressed and are stable. This suggests that these residues probably do not play a vital structural role for the protein. The next task was to assess which of these residues were important for catalysis using the *in vitro* radioassay developed in Chapter 2. These assays revealed that the E102A, E112A, E118A, R586A and Y591A mutations had no significant impact on *C*-mannosyltransferase activity (Figure 4.2E). However, the Y75F, H375A, and E579A mutations did result in a reduction in *C*-mannosyltransferase activity, suggesting that they may play a role in substrate recognition and/or catalysis (Figure 4.2F).

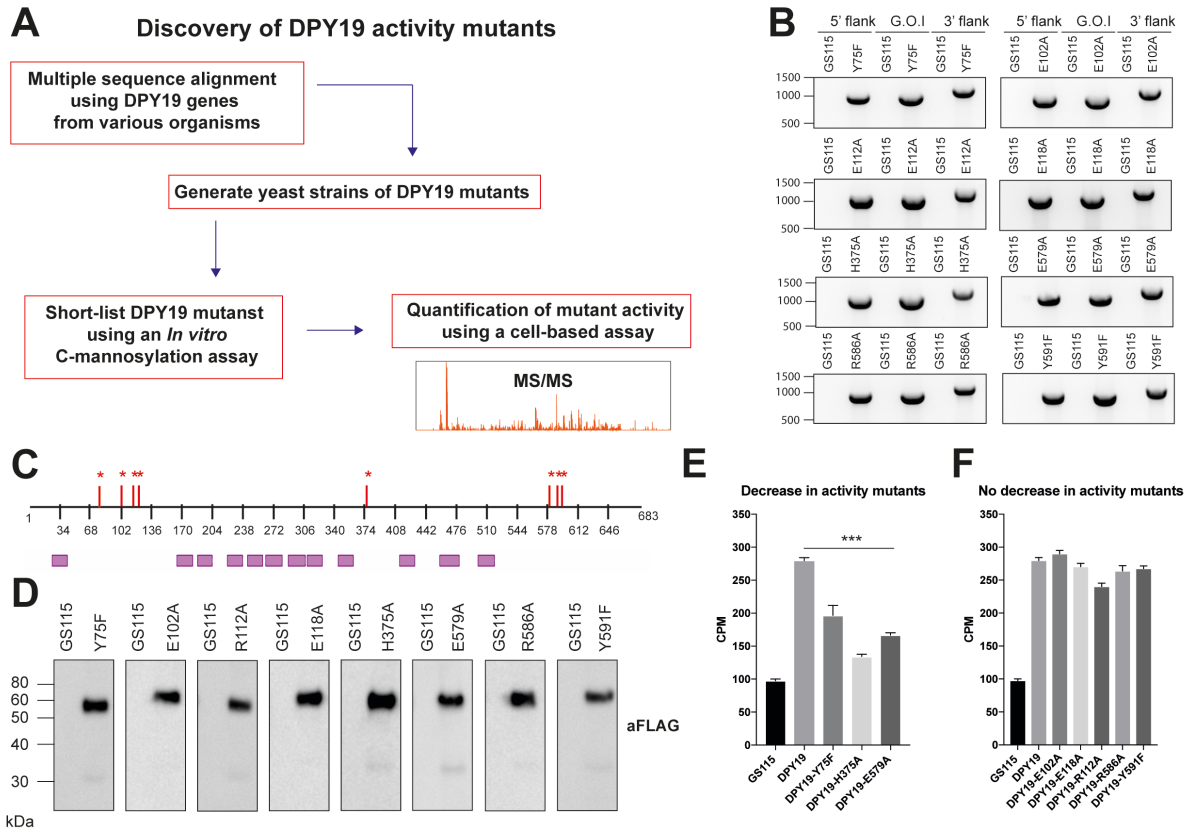


Figure 4. 2: Short-listing DPY19 activity mutants

a) Strategy for the identification of DPY19 mutants with a decrease in activity based on multiple sequence alignment identifying conserved residues.

b) Agarose gel of the polymerase chain reaction confirming integration at the 5' flank, DPY19 gene, and 3' flank for eight DPY19 mutants

c) Schematic showing the location of the residues found the be conserved in DPY19 genes from various organisms. The blue segments represent transmembrane regions of DPY19.

d) Western blots for the eight different DPY19 mutants. An anti-flag antibody was used for the western blots.
e) *In vitro C*-mannosylation assay showing DPY19 mutants with a significant decrease in activity. All data is in triplicates. *** p-value <0.001
f) *In vitro C*-mannosylation assay showing DPY19 mutants with no significant decrease in activity. All data is in triplicates.

The activity of these mutants within the cell were then assessed by integrating a cassette encoding inducible IL12B, which is mannosylated on tryptophan and serves as a useful reporter protein that can be purified and analysed using tandem mass spectrometry (Figure 4.3A). No difference in the level of IL12B expression was observed in these three mutants (Figure 4.3B). Figure 4.3C shows the summary of the level of total *C*-mannosylation for IL12B produced in GS115, DPY19-WT, DPY19-Y75F, DPY19-H375A, and DPY19-E579A. There was a slight reduction in *C*-mannosylation in both the DPY19-Y75F and DPY19-H375A mutants indicating that these residues might play a minor role in substrate binding and/or catalysis (Figure 4.3C). However, for the DPY19-E579A mutant, there was a near-complete loss of C-mannosyltransferase activity indicating that this residue is crucial for catalysis (Figure 4.3C).

Upon close investigation of the different glycoforms of IL12B, we see differences in *C*-linked and *O*-linked glycosylation in these different DPY19 mutants. For IL12B produced in DPY19-WT, 92% of IL12B is doubly modified with *C*-mannosylation, with very little *O*-linked glycosylation (Figure 4.3E). For IL12B produced in DPY19-Y75F, we get 68% *C*-linked glycosylation (*C*-mannosylation) but we get an increase in triply modified (14%) (2 x *C*-linked, 1 x *O*-linked) glycosylation (Figure 4.3F). For DPY19-H375A, we get a very similar glycosylation pattern as for DPY19-Y75F, but a slight increase in the singly glycosylated (*C*-linked) glycoform (Figure 4.3G). However, for IL12B produced in the DPY19-E579A mutant we see a dramatic loss of *C*-linked (*C*-mannosylation) glycosylation (Figure 4.3H). Unmodified IL12B accounts for 65% of all of IL12B, 10% singly glycosylated *O*-linked glycosylation and 3% *C*- and *O*- linked glycosylation. We get 4% triply *O*-linked (3 x *O*-linked) glycosylation and 9% quad *O*-linked (4 x *O*-linked) glycosylation (Figure 4.3H). The severely diminished *C*-linked (*C*-mannosylation) glycosylation means that E579 is crucial for *C*-mannosyltransferase activity of DPY19. Figure 4.3I shows a summary of the different glycoforms of IL12B produced in the various strains. IL12B-E579A has diminished singly and

doubly glycosylated states as compared to IL12B-WT, while IL12B-Y75F and IL12B-H375A have slight reductions in their doubly glycosylated glycoforms.

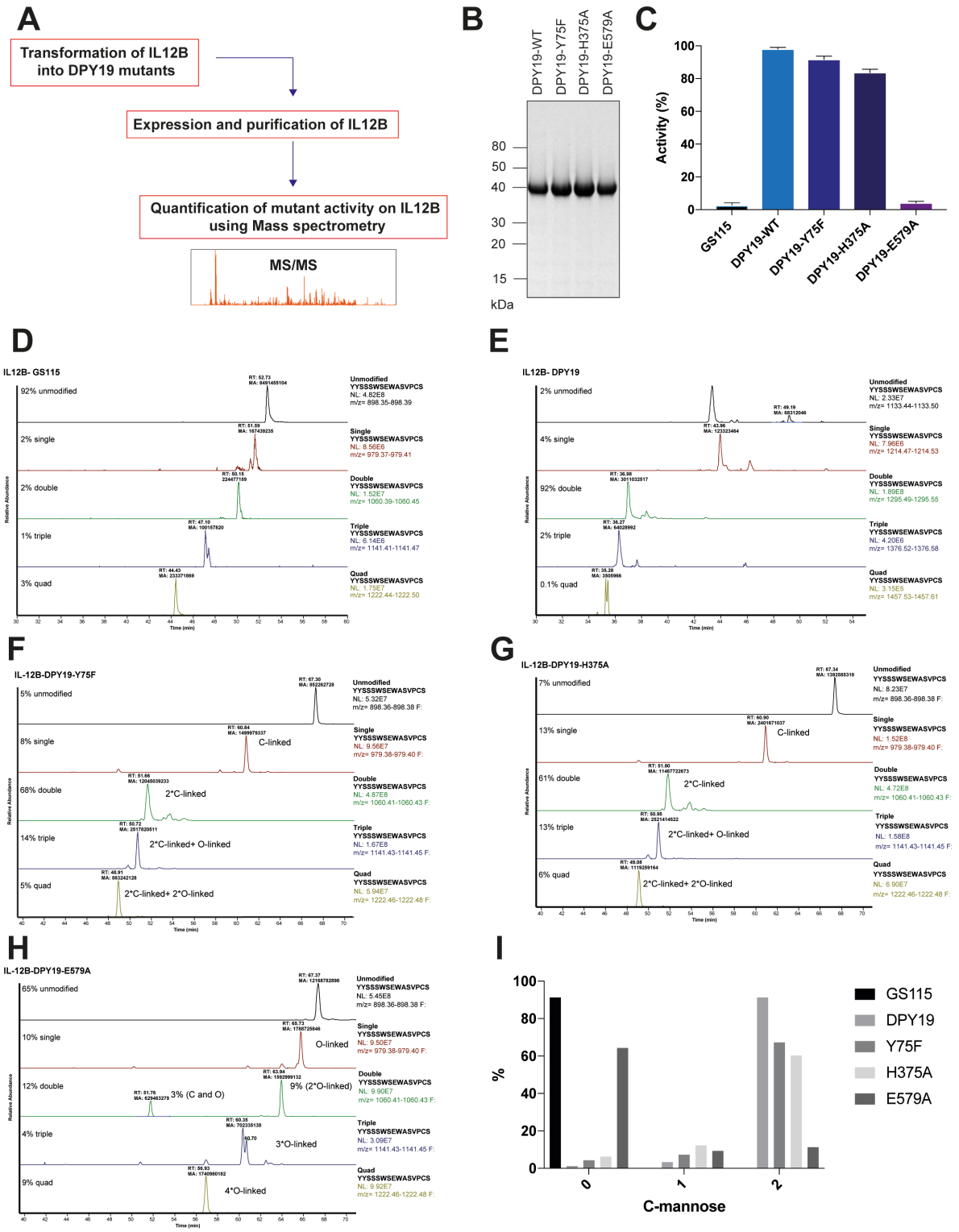


Figure 4. 3: A key conserved residue amongst the DPY19 family is crucial for catalysis

a) Strategy for conducting the cell-based assay for DPY19 mutant activity.

b) SDS-PAGE for DPY19-WT, DPY19-Y75F, DPY19-H375A and DPY19-E579A.
c) Summary of the cell-based total *C*-mannosyltransferase activity on the three DPY19 mutants.
d) Extracted ion chromatogram for IL12B produced in the *GS115* strain showing the different glycoforms. *C*-linked refers to *C*-mannosylation whereas *O*-linked refers to *O*-glycosylation.
e) Extracted ion chromatogram for IL12B produced in the DPY19-WT strain showing the different glycoforms. *C*-linked refers to *C*-mannosylation whereas *O*-linked refers to *O*-glycosylation.
f) Extracted ion chromatogram for IL12B produced in the DPY19-Y75F strain showing the different glycoforms. *C*-linked refers to *C*-mannosylation whereas *O*-linked refers to *O*-glycosylation.
g) Extracted ion chromatogram for IL12B produced in the DPY19-H375A strain showing the different glycoforms. *C*-linked refers to *C*-mannosylation whereas *O*-linked refers to *O*-glycosylation.
h) Extracted ion chromatogram for IL12B produced in the DPY19-E579A strain showing the different glycoforms. *C*-linked refers to *C*-mannosylation whereas *O*-linked refers to *O*-glycosylation.
i) Summary of the different *C*-linked glycoforms for IL12B produced in the *GS115*, DPY19-WT, DPY19-Y75F, DPY19-H375A and DPY19-E579A strains.

4.3 - Discussion

The DPY19 family is a highly conserved protein family that has only recently been discovered, and of the four *dpy19* homologues in human, there is only evidence for activity for two of them [71, 74]. It may be that these other two homologues are inactive or may have different substrates which have not been identified yet. Because of this, not very much is known about this family and there is no crystal structure showing what the enzyme looks like. Considering that DPY19 homologues possess 12-13 transmembrane helices, crystallising or even purifying these complex enzymes is very difficult. As demonstrated in chapter 2 of this thesis, the *C. elegans* DPY19 enzyme is unstable after 24h at 4 °C, and does not even survive a freeze thaw cycle, which makes it very difficult to purify and perform experiments on at an *in vitro* level. We were also not able to express the DPY19L1 and DPY19L3 human homologues of DPY19

in *P. pastoris*. Given that the *C. elegans* DPY19 is unstable due to temperature, it may be that DPY19L1 and DPY19L3 are also unstable proteins at temperatures of 30 °C, the temperature at which *P. pastoris* grows. It also may be that these proteins require specific chaperones to fold, which may only be found in specific higher eukaryotic organisms.

The data in this chapter has shown that among the DPY19 protein family there are a number of conserved amino acid residues, of which one of them E579, is crucial for catalysis. This is a very interesting find because of the four DPY19 homologues in mammals, E579 is only found in DPY19L1-3, and is absent in DPY19L4. We know that DPY19L1 and DPY19L3 are active *C*-mannosyltransferases which might need this specific residue for their catalysis. DPY19L2 and DPY19L4 are not active *C*-mannosyltransferases on the substrate proteins that they have been tested on [71] [74]. It is not clear why DPY19L2 has this residue given that all attempts to show its activity as a *C*-mannosyltransferase have failed thus far on the substrates tested, but this does not rule out that it may be active on other substrates. Perhaps there are other residues that are involved in catalysis. Perhaps these other residues are not conserved, because their importance lay in their properties and can be substituted by similar residues, for instance basic residues, acidic residues or charged residues.

This might give some indication that for *dpy19l4*, it has evolutionarily diverged from the *C*-mannosyltransferase function and is possibly providing a different function to *C*-mannosylation. Considering that *dpy19l4* does not contain the conserved E579 amino acid that is required for catalysis in the *C. elegans dpy19*, it may be that *dpy19l4* has evolved to perform a different function, whether it is enzymatic or structural. It may still be a glycosyltransferase, but it may not be installing a mannose onto a tryptophan. It may be installing another carbohydrate molecule, such as a glucose, galactose or fucose onto another amino acid. It may be that this protein may have undergone mutations to change it into a structural protein, such as *dpy19l2* [120]. It may be that it is a receptor, or may have even become a transporter.

It is interesting how the homologues among the *dpy19* gene family are so different at the sequence level as compared to each other. We know that an ancestral *dpy19* gene gave rise to *dpy19l1* and *dpy19l3*, and *dpy191* gave rise to *dpy19l2* in mammals and vertebrates and *dpy19l3* gave rise to *dpy19l4* [114]. However, the striking difference even among active *C*-mannosyltransferases, *dpy19l1* and *dpy19l3*, suggests that delineating their mechanism of *C*-mannosylation may reveal that these two enzymes have different target proteins. *Dpy19l3* and

dpy19l4 have only a 43% similarity supporting the idea that *dpy19l4* still has function but its function is different to the other homologues.

It would be interesting to mutate the E579 residue in mammalian DPY19L1 and DPY19L3, since they are both *C*-mannosyltransferases and investigate how this impacts their activity. We also provide some insight into the topology of the DPY19 family and show that most of the conserved residues are found in soluble non-transmembrane regions indicating that these regions might be where the active site of this enzyme lies. Ultimately, crystallising the *C. elegans* DPY19 protein and the mammalian DPY19L1 and DPY19L3 proteins would give us a complete picture of how this enzyme family works and delineate the intricate differences between DPY19L1 and DPY19L3.

4.4 - Conclusions

Our investigation into DPY19 through multiple sequence alignments led to the discovery of highly conserved amino acid residues among the DPY19 family. Mutants were made for each conserved amino acid for the *C. elegans dpy19* gene and their activities were tested *in vitro* and inside cells and it was discovered that E579 is crucial for catalysis.

Chapter 5

Generation of antibodies against C^2-α-D-mannosyl- tryptophan for the identification of *C*-mannosylated proteins

5.1 - Introduction

Since the discovery of *C*-mannosylation in 1994 on RNaseII from human urine, only approximately 30 other proteins have been demonstrated to possess this modification [8] and the tools for detecting the modification have remained unchanged. Mass spectrometry is the main method for detecting hexose-modified tryptophan and, while it provides valuable localisation and occupancy information, it is expensive, medium-to-low throughput, and cannot definitely assign the attached hexose as an α-D-mannoside. A simple way to overcome these limitations would be to generate monoclonal antibodies against C^2-α-D-mannosyl-tryptophan. These antibodies, together with modern immunochemistry methods, could be used to identify a large number of the *C*-mannosylated proteins predicted to exist within the proteome, to validate DPY19 knockout cell lines and to aid in the development of small molecule inhibitors of DPY19. While no monoclonal antibodies have been described to date, an affinity purified polyclonal serum against C^2-α-D-mannosyl-tryptophan has been reported [165], though it is no longer available to the scientific community. This illustrates the feasibility and need for the generation of monoclonal antibodies to C^2-α-D-mannosyl-tryptophan.

In this chapter, I demonstrate how my engineered DPY19 yeast strain can be used to prepare a peptide bearing C^2-α-D-mannosyl-tryptophan, and how this peptide was used to raise monoclonal antibodies that are specific for the modification. The selectivity of these antibodies was evaluated against a panel of *C*-mannosylated proteins using Western blots and their affinities for antigen were determined using SPR. Sequencing the complementarity-determining regions (CDRs) of all monoclonal antibodies identified from two mice revealed immune-convergence for the recognition of C^2-α-D-mannosyl-tryptophan. The F_{ab} domain of one antibody was co-crystallised with the mannosylated peptide antigen and x-ray diffraction techniques provided molecular-level insights into the interactions between these mAbs and C^2-

α-D-mannosyl-tryptophan. Finally, these antibodies were used to provide evidence of new *C*-mannosylated proteins in fetal calf serum, human serum and *Toxoplasma gondii* tachyzoites.

5.2 - Results

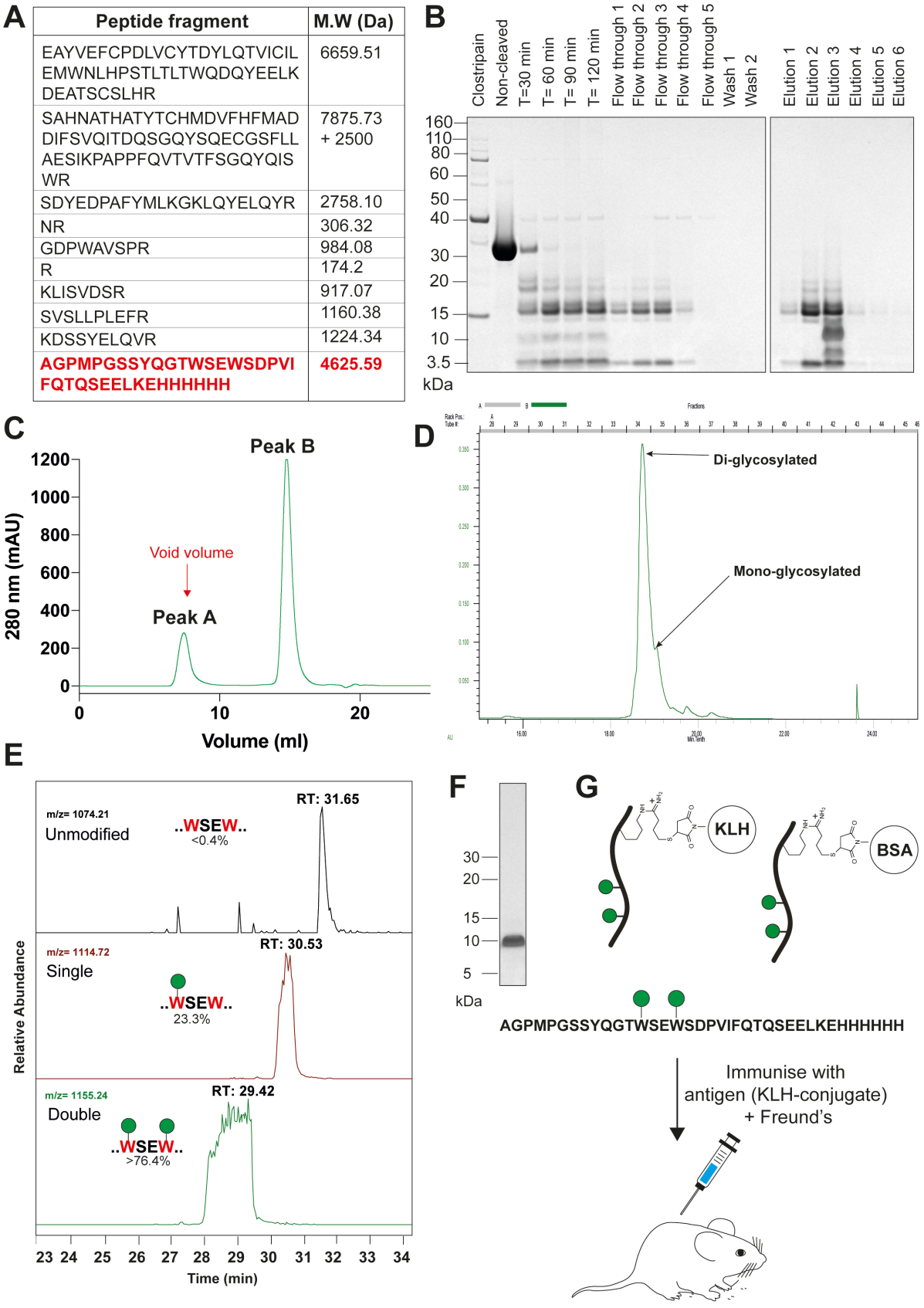
5.2.1 - Generation of a *C*-mannosylated antigen using the DPY19 yeast strain

Synthesising C^2-α-D-mannosyl-tryptophan and incorporating this glyco-amino acid into peptide antigens is a demanding process [166-168] and several attempts to emulate this chemistry in our laboratory have been unsuccessful. Initial attempts to raise antibodies against C^2-α-D-mannosyl-tryptophan involved immunising mice with recombinant UNC5 produced in my yeast strain harbouring *C. elegans* DPY19, which had been shown by mass spectrometry to be extensively glycosylated (see Chapter 2). With three mannosylation sites, it was reasoned that UNC5 would provide a diverse array of epitopes and encourage an immune response to C^2-α-D-mannosyl-tryptophan. However, the serum from these mice failed to demonstrate a strong antibody titre against UNC5, as determined by ELISA assays and Western blots. Clearly this small disulphide rich protein was not immunogenic enough for this experiment.

An alternative strategy for antigen generation was adopted, which involved the proteolysis of recombinant protein from the DPY19 yeast strain, purification of the mannosylated peptide, and conjugation to an immunogenic carrier protein. To this end, the ectodomain of human IL21R produced in our DPY19 strain (see Chapter 2) was digested by clostripain (Endoproteinase-Arg-C) which cleaves the amide bond following arginine residues [169]. Clostripain cuts at nine sites in this IL21R construct to give a total of 10 fragments with molecular weights up to ≈8 kDa (Figure 5.1A). Our peptide of interest, AGPMPGSSYQGTWSEWSDPVIFQTQSEELKEHHHHHH, has two C^2-α-D-mannosyl-tryptophan residues and a C-terminal hexahistidine (His_6) tag for purification using immobilized metal affinity chromatography (IMAC). The optimal conditions for clostripain digestion were determined to be 20 μg of clostripain per milligram of IL21R at 25°C for 2 h (Figure 5.1B). After digesting 40 mg of recombinant IL21R with clostripain, the peptide of interest was captured using IMAC (Figure 5.1B). This IMAC fraction contained a substantial number of impurities, therefore we decided to further purify the peptide using size exclusion chromatography (SEC). The SEC chromatogram gave us two major peaks. The first peak

eluted around 7.5 ml, which is where we would expect the void volume: this turned out to possess most of the impurities (Figure 5.1C), as determined by SDS-PAGE. The second major peak at 15 ml was predominantly the mannosylated peptide of interest and was about 80% pure. This sample was purified further by high-performance liquid chromatography (HPLC) (Figure 5.1D) on a C8 semi-prep column. The peptide of interest was eluted using a gradient of 5-100% acetonitrile and 0.1% trifluoroacetic acid (TFA) across 12 column volumes (30 ml) to give a peak at 18 to 20 min (Figure 5.1D). LC-MS/MS confirmed that this material was the peptide AGPMPGSSYQGTWSEWSDPVIFQTQSEELKEHHHHHH, and that the peptide contained zero, one and two C^2-α-D-mannosyl-tryptophans (Figure 5.1E). The majority of the peptide (>76%) was doubly modified and about 23% was singly modified. An SDS-PAGE of this material provided further evidence of its purity (Figure 5.1F).

This peptide, which possesses a lysine side chain amino group and an N-terminal amine, was activated with Traut's reagent to install a sulfhydryl group. This was conjugated to maleimide-activated keyhole limpet hemocyanin (KLH) and maleimide-activated bovine serum albumin (BSA) using the manufacturer's protocols (Figure 5.1G). KLH is a carrier protein known to enhance the immunogenicity of molecules that are not generally immunogenic [170]. KLH is a very large macromolecular complex with many different epitopes and is xenogeneic to mammalian organisms, which significantly increases its immunogenic response [171]. The peptide-BSA conjugate was prepared for screening antibodies specific to the peptide but not the KLH carrier.



Peptide fragment	M.W (Da)
EAYVEFCPDLVCYTDYLQTVICIL EMWNLHPSTLTLTWQDQYEELK DEATSCSLHR	6659.51
SAHNATHATYTCHMDVFHFMAD DIFSVQITDQSGQYSQECGSFLL AESIKPAPPFQVTVTFSGQYQIS WR	7875.73 + 2500
SDYEDPAFYMLKGKLQYELQYR	2758.10
NR	306.32
GDPWAVSPR	984.08
R	174.2
KLISVDSR	917.07
SVSLLPLEFR	1160.38
KDSSYELQVR	1224.34
AGPMPGSSYQGTWSEWSDPVI FQTQSEELKEHHHHHH	**4625.59**

Figure 5. 1: Generation of an immunogenic *C*-mannosylated antigen.

a) Table listing the IL21R peptide fragments obtained from digestion with clostripain.

b) SDS-PAGE of fractions from the clostripain-digestion and IMAC purification of IL21R. The band at 10 kDa is the peptide of interest.
c) SEC chromatogram showing the separation of peptides from 'elution 3' of Figure 5.1B. Peak A is the void volume and contains impure peptide aggregates. Peak B is the peptide of interest.
d) HPLC chromatogram for purification of Peak B from Figure 5.1C. Mono-mannosylated peptide is the major peak and di-mannosylated peptide is the shoulder to the right of this peak.
e) Extracted ion chromatogram (EIC) of fraction 34 and 35 from Figure 5.1D showing the relative abundance of peptide glycoforms.
f) SDS-PAGE of the purified peptide.
g) The peptide was activated with Traut's reagent, conjugated to maleimide-activated KLH and BSA. The KLH conjugate was used to immunise mice with Freund's complete adjuvant. Two boosters with Freund's incomplete adjuvant were later provided.

5.2.2 – The peptide-KLH conjugate provides a strong immune response in mice

Four mice were immunised with the peptide-KLH conjugate and Freund's complete adjuvant. Each were given three boosters of peptide-KLH conjugate and Freund's incomplete adjuvant over the course of three months. Serum was collected from these animals one week after the last booster and evaluated by ELISA using the BSA-conjugated peptide (Figure 5.2A): all animals demonstrated a high antibody titre against the peptide of interest. We then assessed the serum of these mice through Western blot against a panel of proteins with and without C^2-α-D-mannosyl-tryptophan (Figure 5.2B). We decided to use IL21R, UNC5 and IL12B with C^2-α-D-mannosyl-tryptophan and IL12B without C^2-α-D-mannosyl-tryptophan. We were looking for mice with serum that reacted to proteins other than IL21R only in the presence of C^2-α-D-mannosyl-tryptophan. Mouse #1636 and #1637 gave us this exact result, with strong recognition of IL21R, and bands for IL12B with C^2-α-D-mannosyl-tryptophan but not without it (Figure 5.2B).

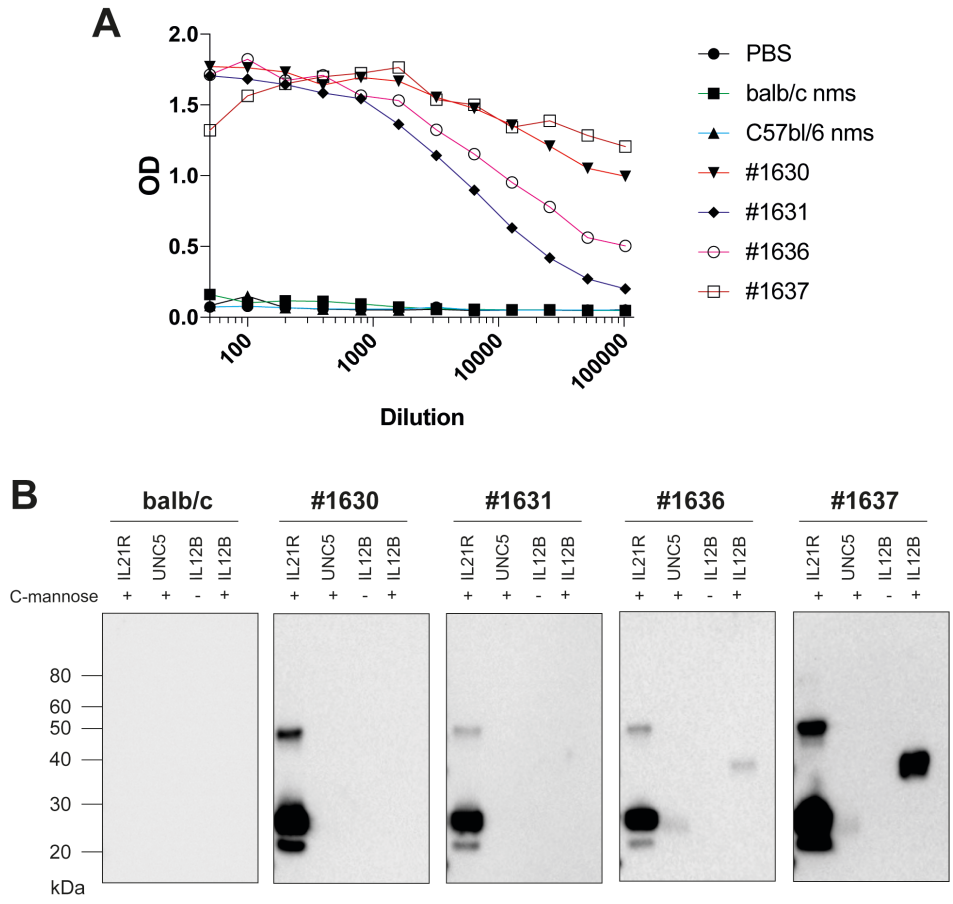


Figure 5. 2: Some mice immunised with the peptide-KLH conjugate produce antibodies that only recognise *C*-mannosylated proteins

a) ELISA titration results showing reactivity of immunised mouse serum to the peptide-BSA conjugate (mouse ID#: 1630, 1631, 1636 and 1637).

b) Western blot results for serum from mouse 1630, 1631, 1636, 1637 as well as a naïve balb/c control. The proteins IL21R, UNC5 and IL12B were produced in yeast with or without DPY19 to provide proteins with and without C^2-α-D-mannosyl tryptophan. Serum was diluted to 1/1000 and secondary anti-mouse antibody was used at 1/10,000.

5.2.3 – Generation of mAbs that recognise C^2-α-D-mannosyl-tryptophan

Hybridomas were generated from mice #1630, #1636 and #1637 using the spleen fusion method (Figure 5.3A) [172-174]. After 2 weeks, hybridoma supernatants from these mice were screened by ELISA for their reactivity to the BSA-conjugated peptide, resulting in 48 hits (Figure 5.3B–D). Western blots were conducted on all of these hits using the same protein analytes as before. Five hybridomas produced antibodies that were able to recognise C^2-α-D-mannosyl-tryptophan in different proteins (IL21R and IL12B), 23 were specific for the IL21R

peptide, and 20 did not recognise any of the proteins (Figure 5.3E–G). Mouse #1637 yielded two clones (5G12 and 10E9) (Figure 5.3A) and mouse #1636 provided three clones (9E7, 6G4 and 7G12) (Figure 5.3G). These five hybridoma clones underwent several rounds of serial sub-cloning by limiting dilution and after each round, the supernatant from the sub-clones was assessed for reactivity to C^2-α-D-mannosyl-tryptophan by Western blot. The final five hybridoma clones were used to produce 1 mg of pure antibodies using the protein-A affinity purification method.

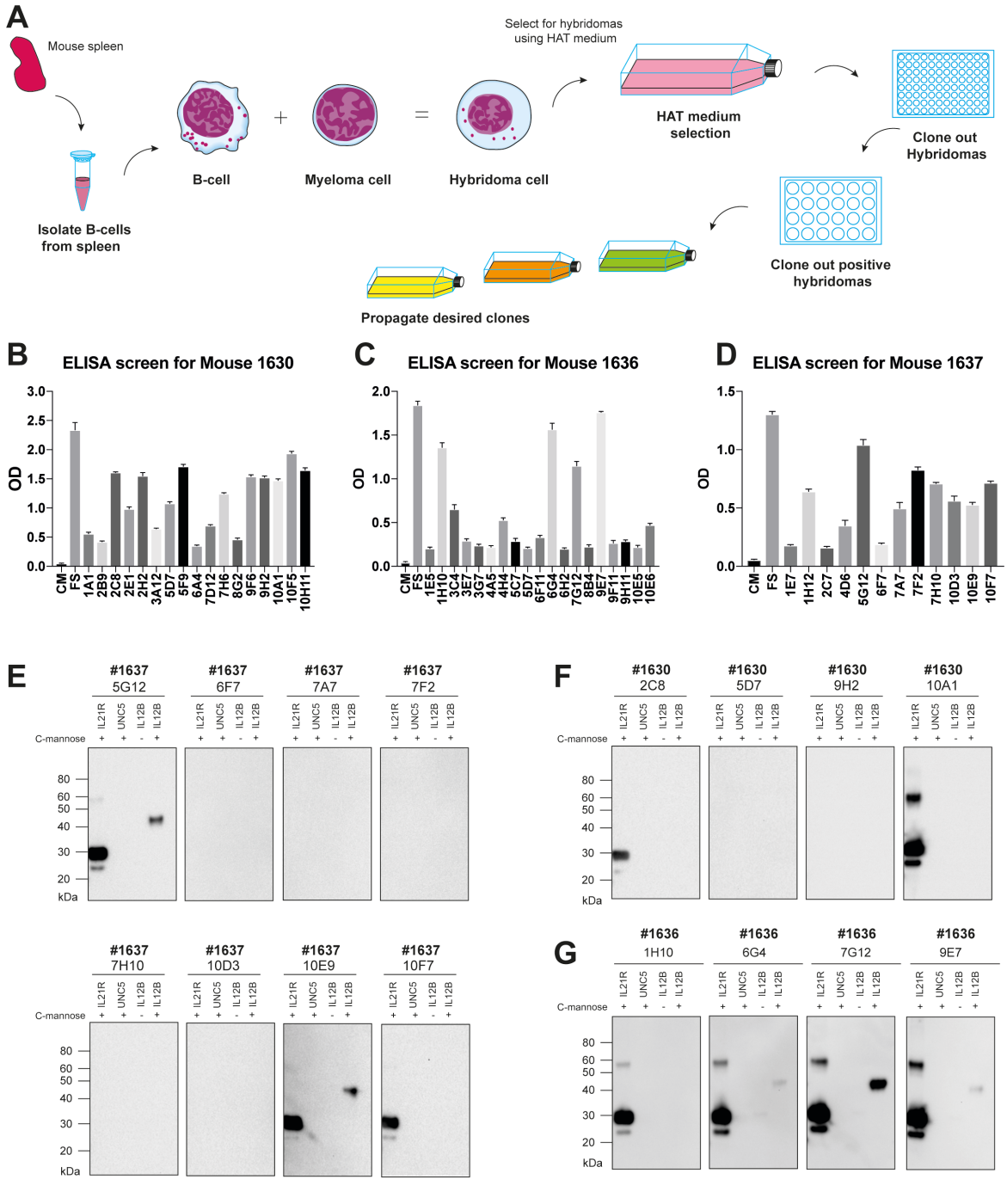


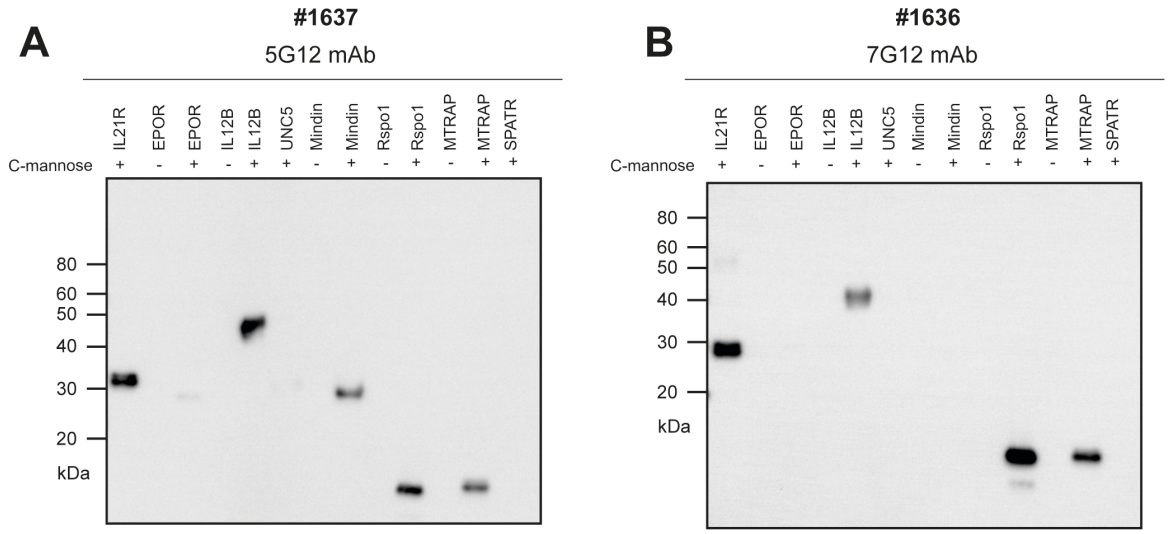
Figure 5. 3: Short-listing hybridomas that recognise C^2-α-D-mannosyl tryptophan

a) Illustration showing how hybridoma clones are produced.
b) ELISA results for reactivity of the peptide-BSA conjugate to hybridoma clones from mouse 1630. Hybridoma supernatant was diluted 1/100 and tested against 100 ng of antigen-BSA conjugate.
c) ELISA results for reactivity of the peptide-BSA conjugate to hybridoma clones from mouse 1636. Hybridoma supernatant was diluted 1/200 and tested against 100 ng of antigen-BSA conjugate
d) ELISA results for reactivity of the peptide-BSA conjugate to hybridoma clones from mouse 1637. Hybridoma supernatant was diluted 1/200 and tested against 100 ng of antigen-BSA conjugate
e) Western blot results for selected positive hybridoma clones from mouse #1637 using IL21R, UNC5 and IL12B with C^2-α-D-mannosyl tryptophan and IL12B without C^2-α-D-mannosyl tryptophan. 100 ng of reporter proteins was used. Supernatant was diluted to 1/50 and secondary anti-mouse antibody was used at 1/10,000.
f) Western blot results for selected positive hybridoma clones from mouse #1630 using IL21R, UNC5 and IL12B with C^2-α-D-mannosyl tryptophan and IL12B without C^2-α-D-mannosyl tryptophan. 100 ng of reporter proteins was used. Supernatant was diluted to 1/50 and secondary anti-mouse antibody was used at 1/10,000.
g) Western blot results for selected positive hybridoma clones from mouse #1636 using IL21R, UNC5 and IL12B with C^2-α-D-mannosyl tryptophan and IL12B without C^2-α-D-mannosyl tryptophan. 100 ng of reporter proteins was used. Supernatant was diluted to 1/50 and secondary anti-mouse antibody was used at 1/10,000.

5.2.4 - Characterisation of five monoclonal antibodies that recognise C^2-α-D-mannosyl-tryptophan

Two monoclonal antibodies from each mouse were screened using Western blots against a panel of reporter proteins produced with and without *C*-mannosylation in the DPY19 yeast strain in order to evaluate the selectivity of these reagents. Although all these reporter proteins have a WX_1X_2W motif, the amino acids X_1 and X_2 can vary markedly, as can the residues either side of this motif. The 5G12 mAb from mouse #1637 was absolutely specific for *C*-mannosylated proteins. It had strong recognition for recombinant IL21R, IL12B and Rspo1 (Figure 5.4A), which possess a Ser-Glu-Trp(Man) tripeptide in their amino acid sequence

(Figure 5.4C). It had slighter weaker recognition for MTRAP and mindin: MTRAP has a Ser-Glu-Trp(Man) tripeptide but has low *C*-mannosylation occupancy, while mindin has a Ser-Ser-Trp(Man) tripeptide in its sequence (Figure 5.4A, C). There is also a very weak band for EPOR and UNC5: EPOR has a Ser-Ala-Trp(Man) tripeptide in its sequence, while UNC5 has Ser-Asp-Trp(Man) tripeptide in its sequence (Figure 5.4A, C). For mouse #1636, mAb 7G12 was evaluated by Western blot and found to have strong recognition for IL21R, Rspo1 and MTRAP (Figure 5.4B). All of these proteins have a Ser-Glu-Trp(Man)-Ser tetrapeptide (Figure 5.4C). It also recognises IL12B, which has a Ser-Glu-Trp(Man)-Ala tetrapeptide (Figure 5.4B, C). Clearly both of these mAbs recognise proteins with C^2-α-D-mannosyl-tryptophan, though neither could recognise the modification in all protein contexts. The mAbs from both mice had a strong preference for peptides containing the Ser-Glu-Trp(Man) motif (Figure 5.4C).



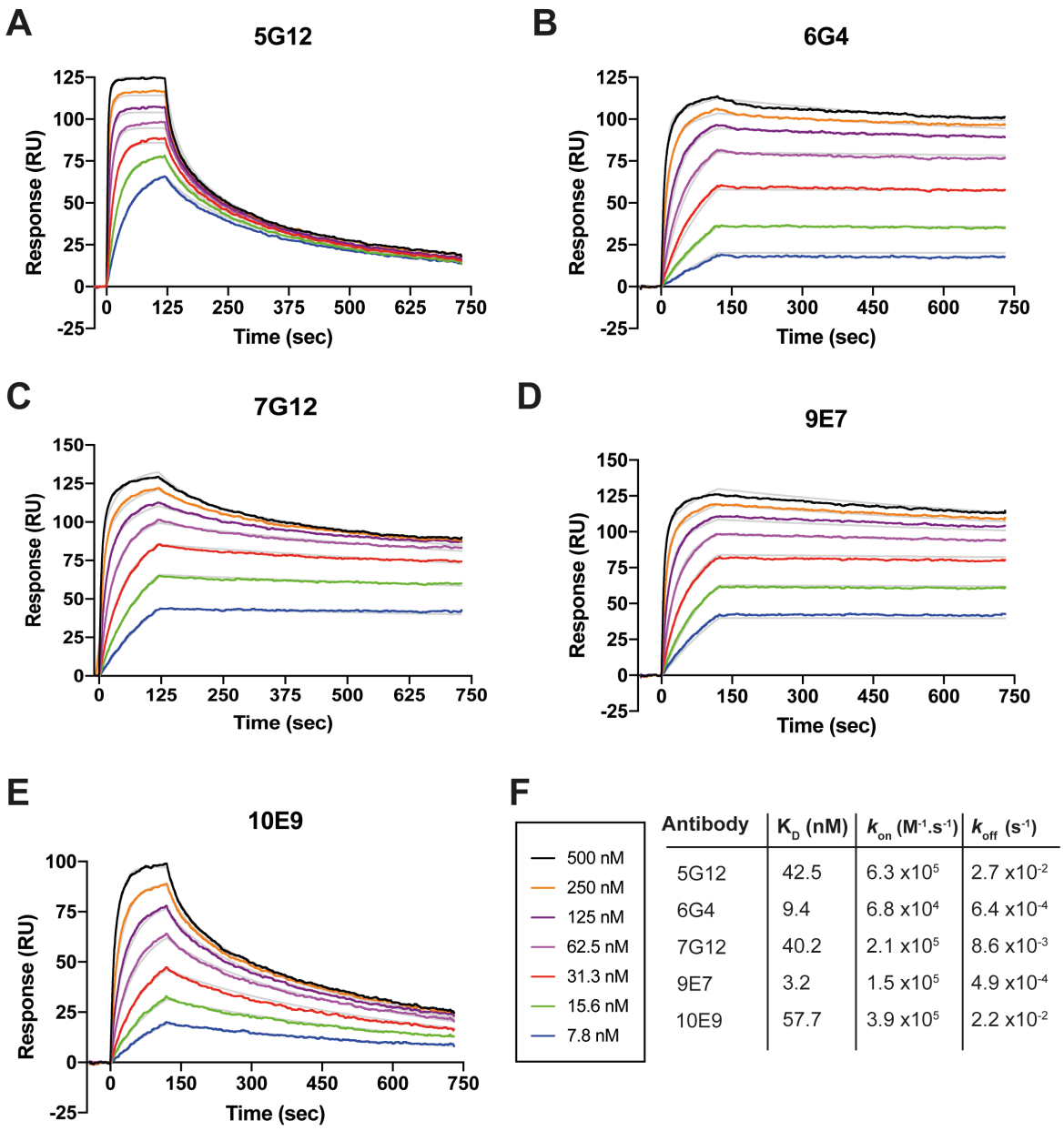
C

Family	Protein	Sequence
Cytokine receptor	IL21R	---YQGTWS**EW**SDPV
	EPOR	---FGGFWSAWSEPV
Other	IL12B	---YSSSWS**EW**ASVP
TSR	UNC5	LDGGWSSWSDWSACS
	Mindin	---EVSLWSSWGLCG
	Rspo1	---EMS**EW**SPWGPCS
	MTRAP1	---TCD**EW**S**EW**SACT
	SPATR	---SYTPWSDWSACK

Figure 5. 4: Antibody epitope mapping using reporter proteins with and without C^2-α-D-mannosyl tryptophan

a) Western blot of reporter proteins with and without C^2-α-D-mannosyl tryptophan mapping out selectivity of 5G12 from mouse #1637. Proteins with a C^2-α-D-mannosyl tryptophan include: IL21R, EPOR, IL12B, UNC5, mindin, Rspo1, MTRAP and SPATR. Proteins without C^2-α-D-mannosyl tryptophan include: EPOR, IL12B, mindin, Rspo1, and MTRAP. 100 ng of each reporter proteins was used. Purified 5G12 was diluted to 1/2000 and secondary anti-mouse antibody was used at 1/10,000.
b) Western blot of reporter proteins with and without C^2-α-D-mannosyl tryptophan mapping out selectivity of 7G12 from mouse #1636. Proteins with a C^2-α-D-mannosyl tryptophan include: IL21R, EPOR, IL12B, UNC5, mindin, Rspo1, MTRAP and SPATR. Proteins without a C^2-α-D-mannosyl tryptophan include: EPOR, IL12B, mindin, Rspo1, and MTRAP. 100 ng of each reporter proteins was used. Purified 7G12 was diluted to 1/2000 and secondary anti-mouse antibody was used at 1/10,000.
c) Table of reporter proteins along with the local amino acid sequence near the WXXW motif. Proteins highlighted in yellow are proteins with strong recognition by 5G12 and 7G12. EW is highlighted in bold because it shows the common sequence that is recognised by the two antibodies.

Next, surface plasmon resonance (SPR) was used to determine the binding affinities of each mAb for the AGPMPGSSYQGTWSEWSDPVIFQTQSEELKEHHHHHH peptide bearing two C^2-α-D-mannosyl-tryptophans. The peptide, which possesses amino groups at the N-terminus and on the lysine side chain, was immobilised to a carboxymethyl cellulose-coated SPR chip by EDC coupling. Each of the mAbs were flowed across the chip at a range of concentrations (500 nM to 7.8 nM) to provide sensorgram response curves that could be fit to a bivalent binding model. This provided association rates (k_{on}) and dissociation rates (k_{off}) for peptide binding to each mAb, which enabled calculation of the dissociation constants (K_D) for each of the five mAbs. The mAbs 5G12 and 10E9 from mouse #1637, had dissociation constants (K_D) of 42 and 58 nM, respectively (Figure 5.5A, E, F). These K_D values were 5–10 fold higher than mAbs 9E7, 6G4 and 7G12 from mouse #1636, which had K_D values of 3.2, 9.4, and 40 nM, respectively (Figure 5.5B-D, F). Differences in these K_D values was mostly driven by variations in the k_{off} rates, as all the mAbs had similar k_{on} rates (Figure 5.5F): both 5G12 and 10E9 had substantially greater k_{off} rates than 9E7, 6G4 and 7G12. The mAbs with the slowest off rates were 9E7 and 6G4, making them the most appropriate for future applications in glycoprotein discovery through immunoprecipitation and affinity purification for mass spectrometry.



Antibody	K_D (nM)	k_{on} ($M^{-1}.s^{-1}$)	k_{off} (s^{-1})
5G12	42.5	6.3×10^{5}	2.7×10^{-2}
6G4	9.4	6.8×10^{4}	6.4×10^{-4}
7G12	40.2	2.1×10^{5}	8.6×10^{-3}
9E7	3.2	1.5×10^{5}	4.9×10^{-4}
10E9	57.7	3.9×10^{5}	2.2×10^{-2}

Figure 5. 5: SPR determination of mAb-peptide binding kinetics.

a) Sensorgram of 5G12 mAb binding to *C*-mannosylated peptide.

b) Sensorgram of 6G4 mAb binding to *C*-mannosylated peptide.

c) Sensorgram of 7G12 mAb binding to *C*-mannosylated peptide.

d) Sensorgram of 9E7 mAb binding to *C*-mannosylated peptide.

e) Sensorgram of 10E9 mAb binding to *C*-mannosylated peptide.

f) Table of binding parameters

5.2.5 - Sequencing the mAbs that bind C^2-α-D-mannosyl-tryptophan

To gain some insights into how unique these antibodies might be, the sequence of the complementarity determining regions (CDRs) was determined using a commercial hybridoma sequencing service (Genscript). Antigen recognition is predominantly determined by the CDRs of both light and heavy chains of the IgG (Figure 5.6A). Remarkably, there was minimal diversity observed amongst the mAbs at a sequence level, even between those originating from different animals (Figure 5.6B-C). In the light chain region, all of the antibodies had identical amino acid sequences except for 7G12, which differed substantially in CDR1 (Figure 5.6B). For the heavy chain, most of the variability came from CDR3, and mAbs from mouse #1637 had almost identical sequences, while mAbs from mouse #1636 had very similar sequences (Figure 5.6C). All of these mAbs appear to utilise a heavy chain derived from the IGKJ1*01 F allele, and four of the five possess a near identical light chain derived from the IGKV1-110*01 F allele, while 7G12 utilised the similar IGKV3-12*01 F allele. Clearly, this combination of heavy and light chains provided an excellent scaffold for the development of high-affinity C^2-α-D-mannosyl-tryptophan antibodies. The most notable difference between mAbs from the two mice lay in CDR3 of the heavy chains. It is tempting to speculate that this particular sequence may be responsible for the substantially slower off-rates, and thus lower dissociation constants, observed for mAbs from this animal.

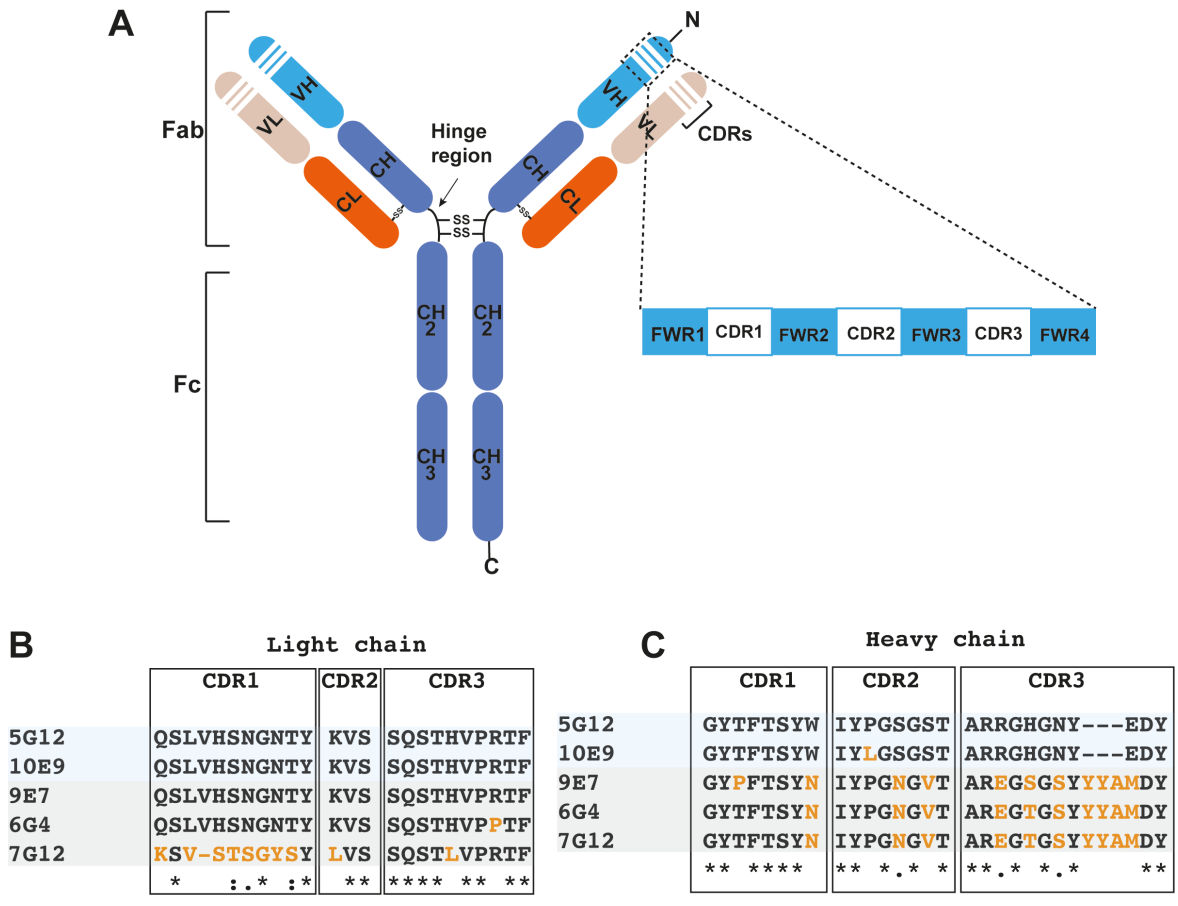


Figure 5. 6: The CDR sequences of antibodies against C^2-α-D-mannosyl tryptophan

a) A cartoon illustrating the structure of a mouse IgG antibody. Variability in the CDRs of the heavy and light chains define what epitopes are recognised by the antibody.

b) Sequence alignment of the light chain CDR regions of the five antibodies from mouse #1636 and #1637.

c) Sequence alignment of the heavy chain CDR regions of the five antibodies from mouse #1636 and #1637.

5.2.6 - Structural insights into antibody recognition of C^2-α-D-mannosyl-tryptophan

Intrigued by the immuno-convergence observed for these antibodies, structural studies were pursued in collaboration with Dr. Michael Jarva, to determine how these heavy and light chain combinations recognised C^2-α-D-mannosyl-tryptophan. The most broadly-reactive mAb, 5G12, was produced from hybridomas and purified using protein A-agarose on a 20-mg scale. This material was digested with papain to yield a F_{ab} and F_c fragments. The F_{ab} fragment was purified by passing the proteolysed mixture through a protein G column to capture the F_c fragment. The flow-through, which contained the F_{ab} fragment, was further purified by SEC

and used in crystallisation trials with the purified IL21R glycopeptide AGPMP GSSYQGTWSEWSDPVIFQTQSEELKEHHHHHH bearing two C^2-α-D-mannosyl-tryptophans. An F_{ab}-peptide crystal appeared after 42 days in a solution containing 10% (v/v) 2-propanol, 20% (w/v) polyethylene glycol 4000, 0.1 M sodium HEPES, pH 7.5, with 150 nL F_{ab}–peptide solution at a 1:1.2 molar ratio at 10 mg/ml (Figure 5.7A). The F_{ab}-peptide complex was collected on a crystallographer's loop and flash frozen with ammonium sulphate cryoprotectant. When shot with X-rays on the MX-2 beamline at the Australian Synchrotron, the crystal provided a diffraction dataset to 1.8 Å (Figure 5.7B).

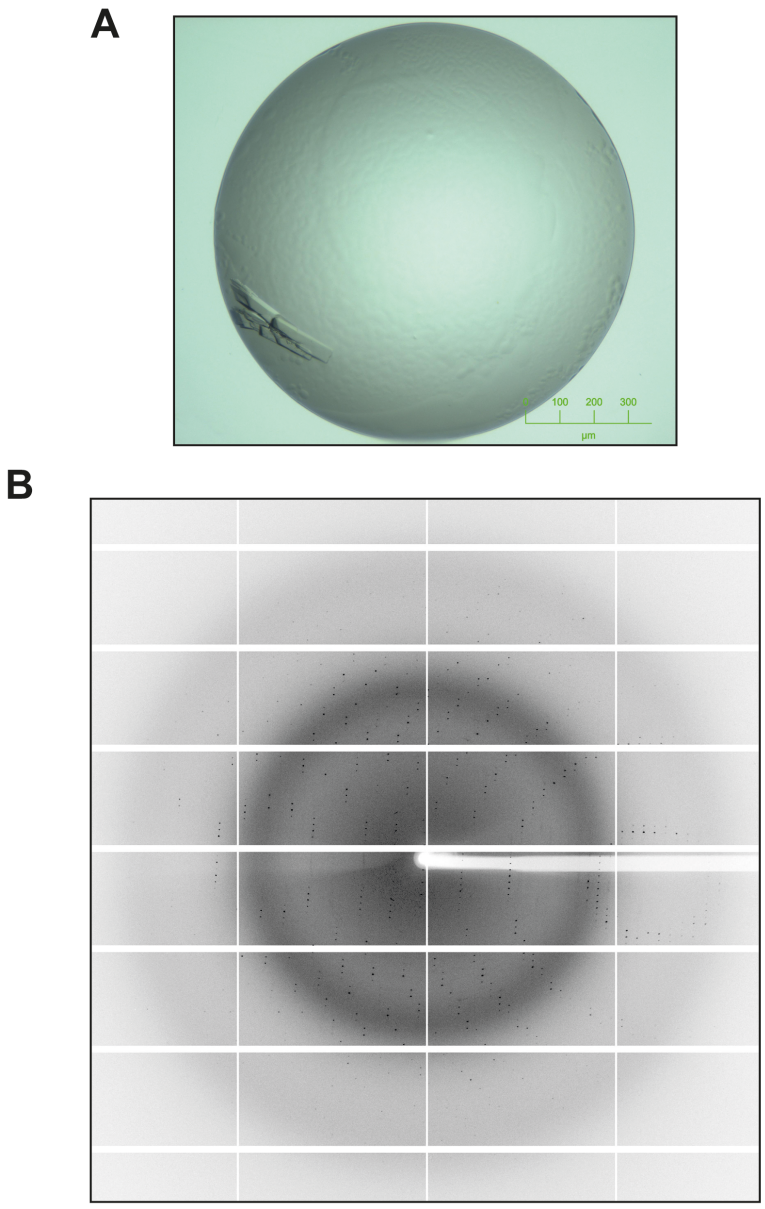


Figure 5. 7: The F_{ab}-peptide complex protein crystal

a) Image of the 5G12 F_{ab}-peptide crystal at day 42. The F_{ab}–peptide crystal was grown over 42 days in sitting drops at 20°C by mixing 150 nl well solution containing 10% (v/v) 2-propanol, 20% (w/v) polyethylene glycol 4000, 0.1 M sodium HEPES, pH 7.5, with 150 nL F_{ab}–peptide solution at a 1:1.2 molar ratio at 10 mg/mL.
b) Representative X-ray diffraction pattern for the F_{ab}-peptide complex protein crystal obtained on the MX-2 beamline at the Australian Synchrotron.

From this diffraction dataset, the structure of the complex was solved in conjunction with Dr. Michael Jarva by molecular replacement using a search model created by removing the variable loops from an IgG_2 F_{ab} (PDB ID: 5MYK). The final structure was refined to a resolution of 1.6 Å with a final R_{work}/R_{free} of 0.1892/0.2207 (Table 5.1).

Table 5. 1: Data collection and refinement statistics

Data collection	
Space group	P 1 2_1 1
No of molecules in AU	3
Cell dimensions	
a, *b*, *c* (Å)	43.23 89.89 59.32
α, β, γ (°)	90.00, 95.94, 90.00
Wavelength (Å)	0.9537
Resolution (Å)*	44.95-1.60 (1.63-1.60)
R_{sym} or R_{merge}*	0.067 (1.282)
R_{pim}*	0.047 (0.927)
$I / \sigma I$*	12.1 (1.0)
CC(1/2)	0.999 (0.393)
Completeness (%)*	97.7 (89.9)
Redundancy*	5.8 (5.1)
Wilson B-factor ($Å^2$)	20.2
Refinement	
Resolution (Å)	44.95-1.60
No. reflections	58,029
R_{work} / R_{free}	0.1892/0.2207
No. non-hydrogen atoms	
FAB	3,398
Peptide	83
Water	370
B-factors	
FAB	33.4
Peptide	48.3
Water	42.0
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angle (°)	0.858
Ramachandran plot (%)	
Favored	97.94

Allowed	2.06
Disallowed	0

*Values in parentheses are for highest-resolution shell.

This high-resolution structure captured the 5G12 F_{ab} in complex with the mannosylated peptide. Positive electron density showed that it bound to a fragment of the peptide corresponding the TWSEWSDP residues (Figure 5.8A, B). We were able to assign positive electron density adjoined to the second tryptophan to a C-linked mannose in its $^1C_4$ pyranose conformation. The CDR3 loops of the mAbs were largely responsible for the interactions between antibody and the mannose and effectively buried the sugar into the cleft of the antibody (Figure 5.8A, B). The C^2-α-D-mannosyl-tryptophan makes the majority of the polar interactions with the F_{ab}. The 4-OH of the mannose makes three hydrogen bonds to $Asp50_H$, $Asn59_H$, and $Arg99_H$ (Figure 5.8C, D). The 5-OH makes an additional three hydrogen bonds, with both the side chain and backbone amide of $Arg101_L$, and a buried water, which in turn is bound to the backbone carbonyls of $Ser96_L$ and $His98_L$ (Figure 5.8C, D). $Arg101_L$ makes a contribution through a hydrogen bonds with the 6-OH of mannose (Figure 5.8C, D). The C^2-α-D-mannosyl-tryptophan is situated such that the aromatic sidechain of tryptophan lies is parallel with the side chain amide of $Asn103_H$ and is close enough to make formal pi-pi stacking interactions (Figure 5.8C, D). The backbone of the tryptophan make two additional hydrogen bonds: the carbonyl interacts with $H31_L$ and the amide with the carbonyl of $Thr97_L$ (Figure 5.8C, D). The glutamate preceding the C^2-α-D-mannosyl-tryptophan creates two notable hydrogen bonds with the sidechain of $S32_L$ and its backbone nitrogen (Figure 5.8C, D), commensurate with these antibody's observed preference for a Glu-Trp(Man) epitope.

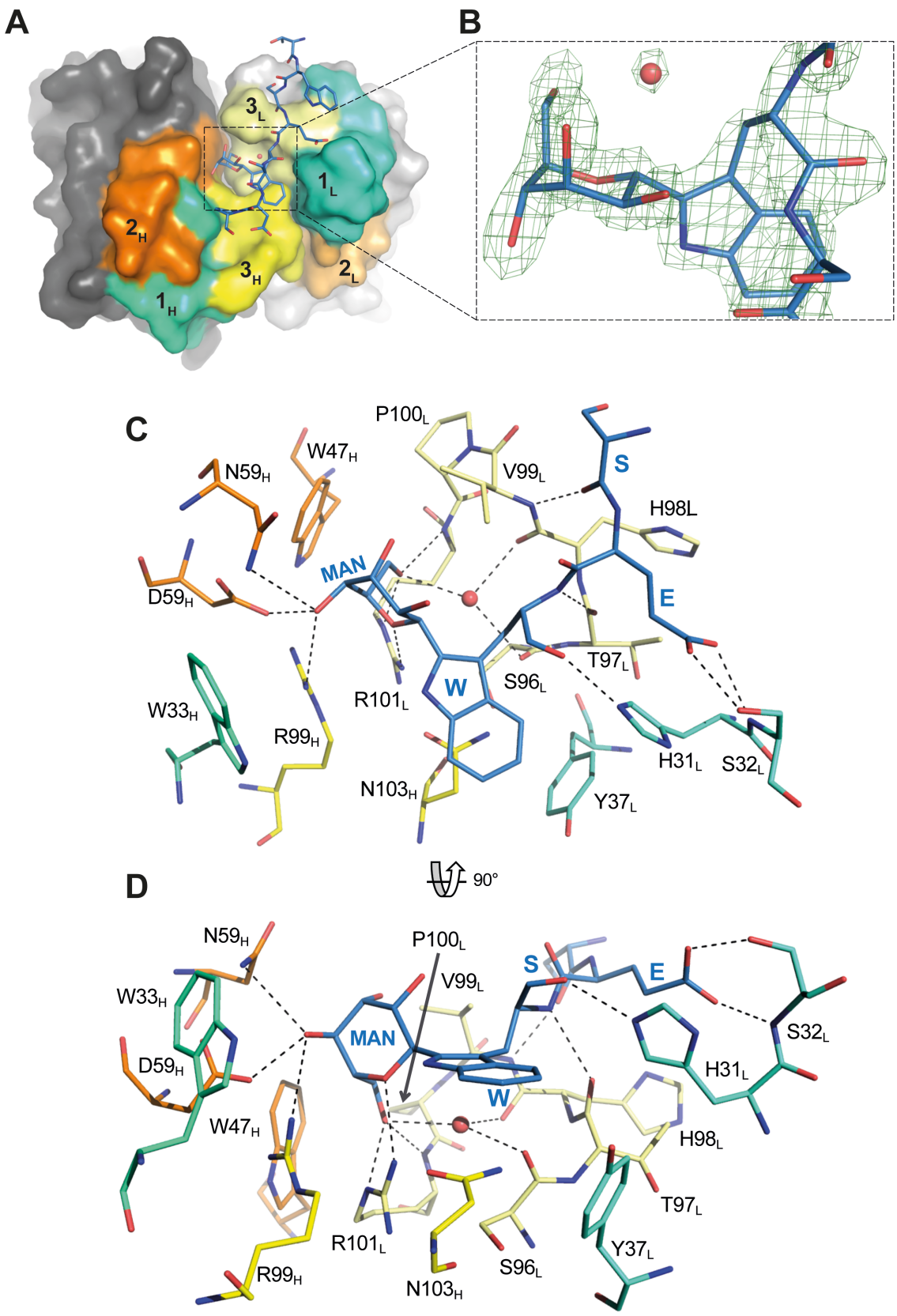


Figure 5. 8: Structural insights into the F_{ab}-peptide complex

a) Top down view of the binding epitope of the F_{ab} showing heavy and light chains in complex with the peptide bearing C^2-α-D-mannosyl tryptophan. The F_{ab} is rendered as a surface model with the variable heavy chain in black (left side), and the variable light chain in grey (right side) with their CDR1-3 loops in cyan, orange, and yellow, respectively.
b) The peptide bearing C^2-α-D-mannosyl tryptophan is depicted using sticks with oxygen in an unbiased Fo-Fc density map at 4-sigma. Water is shown as a red sphere.
c) Residues involved in the peptide binding site with the F_{ab} CDRs colored as above but depicted as sticks. Only the S-E-W residues of the peptide are depicted for clarity.
d) 90-degree rotation of the image in (c).

5.2.7 - Evidence of novel *C*-mannosylated proteins in diverse organisms

The 5G12 antibody, which could recognise C^2-α-D-mannosyl tryptophan in the broadest peptide context, was used to provide evidence for the existence of *C*-mannosylated proteins in diverse organisms. Western blots with samples from fetal calf serum (FCS) (*Bos taurus*), human serum (*Homo sapiens*) and *Toxoplasma gondii* tachyzoite lysate were probed with 5G12 and revealed several bands for each organism (Figure 5.9A-C). For FCS, there were three bands which corresponded molecular weights of 47, 55 and 100 kDa (Figure 5.9A). We saw similar bands in the human serum sample, with an additional band at 25 kDa (Figure 5.9B). These three common bands seen in both the FCS and human serum are quite likely proteins from the complement immune system. Properdin, C6, C7, C8 and C9 are known to be *C*-mannosylated and these are abundant in blood plasma [11, 65]. In the *Toxoplasma gondii* lysate (a generous gift from A/Prof. Chris Tonkin), several bands were observed but the most prominent was a protein of ca. 75 kDa. The most abundant protein in these parasites with a WXXW motif is Microneme protein 2 (MIC2). Fortuitously, a MIC2 knockout cell line was available in the lab of A/Prof Chris Tonkin and, when lysate from this parasite line was probed with 5G12 by Western blot, the strong band at 75 kDa significantly diminished (Figure 5.9C). MIC2 has since been identified as a protein with tryptophan mannosylation [123]. This data reveals that these antibodies have very little 'background' promiscuity and that they can be useful for identifying protein in complex samples. The main limitation of these high quality, well-validated reagents is that they do require the Glu-Trp(Man) dipeptide for reliable detection, though fortunately this is a common motif.

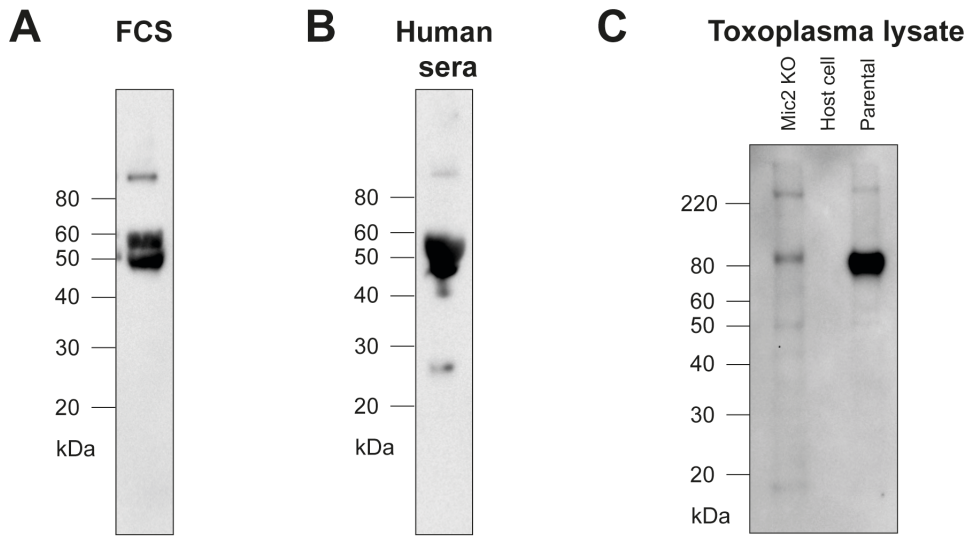


Figure 5. 9: Western blots illustrating the presence of C^2-α-D-mannosyl tryptophan on several proteins in diverse and complex biological samples.

a) Western blot of fetal calf serum using the 5G12 antibody. 60 µg of total protein was used as measured by a Bradford protein assay. Purified 5G12 was used at a dilution of 1/2000 in 3% skim milk in PBS. Secondary anti-mouse antibody was used at 1/10,000 in 3% skim milk in PBS.

b) Western blot of human sera using the 5G12 antibody. 60 µg of total protein was used as measured by a Bradford protein assay. Purified 5G12 was used at a dilution of 1/2000 in 3% skim milk in PBS. Secondary anti-mouse antibody was used at 1/10,000 in 3% skim milk in PBS.

c) Western blot of *Toxoplasma gondii* cell lysate using the 5G12 antibody. 30 µg of total protein was used as measured by a Bradford protein assay. Purified 5G12 was used at a dilution of 1/2000 in 3% skim milk in PBS. Secondary anti-mouse antibody was used at 1/10,000 in 3% skim milk in PBS.

5.3 - Discussion

The data presented here describes the first five monoclonal antibodies capable of recognising peptides bearing C^2-α-D-mannosyl-tryptophan, the generation of which were enabled by the yeast strain engineered to express DPY19. The lack of tools in the field have hindered

investigations into the biological roles of *C*-mannosylation [8]. Previously, Ihara *et al*. described the generation of a polyclonal antibody against C^2-α-D-mannosyl-tryptophan by immunising rabbits with chemically synthesised C^2-α-D-mannosyl-tryptophan [165]. They used this polyclonal affinity-purified antibody to show that Zucker rats, an animal model for diabetes, had increased *C*-mannosylation in proteins as compared to lean rats. The functional reasons and consequences for this observation were never investigated further. However, some of the data presented in this paper are problematic. Western blots using this polyclonal antibody provided inconsistent results in RAW264.7 cells and even showed no significant changes between tissue homogenates from Zucker rats and lean rats. Given how little work was put into characterising this polyclonal antibody, these results must be interpreted carefully. The monoclonal antibodies developed here are much better characterised and could be useful for revisiting the role of *C*-mannosylation in diabetes.

This raises the question of, "what other diseases is *C*-mannosylation involved in?" Considering that there are over 500 proteins with the consensus sequence for *C*-mannosylation and that these proteins are involved myriad cell-signalling pathways, it is likely that changes in *C*-mannosylation are observed in many diseases. The same is probably true for the different stages of development. These antibodies provide us with the means to start tackling these questions.

Sequencing these antibodies led to the interesting observation of immunoconvergence between the antibodies from mouse #1636 and mouse #1637. Of course, the adaptive humoral immune response mediated by B cells has accesses to the same repertoire of VDJ rearrangement in each mouse, but the size of this hypothetical repertoire is vast, which makes convergence on near-identical solutions of interest [175]. At the sequence level, the only real difference for 5G12 and 10E9, which are derived from the same mouse, is in the heavy chain CDR2 at position 53 (L vs P). The noticeable differences in the k_{on} and k_{off} rates for ligand binding by these antibodies suggests that difference in the framework regions (FWR) may also be significant. Antibodies 6G4 and 9E7 from mouse #1636 are almost identical in the CDR regions. The heavy chain CDR1 at position 28 for 9E7 is P but in 6G4 it is T, however, if we look at the crystal structure of 5G12 we can tell that this site is not near the binding site and may not play a significant role in recognition of the ligand. A real difference between these two antibodies can be seen at position 101 in the light chain CDR3, where a R is present in 9E7 and a P is present in 6G4. If we look at the crystal structure of 5G12, the R at that position makes important

interactions between O5 and O6 of the mannose ring, which are not going to be present in 6G4 and may explain its lower affinity relative to 9E7.

The crystal structure of 5G12 F_{ab} in complex with the peptide antigen illustrated the importance of the Glu-Trp(Man) dipeptide for binding. The structure reveals that the light and heavy chain CDR3 regions make most of the interactions with C^2-α-D-mannosyl-tryptophan through extensive hydrogen bonding interactions. The mAb also makes considerable interactions through light chain CDR1 and heavy chain CDR2 with the peptide. The electron density observed for the mannose provides unambiguous assignment of a $^1C_4$ pyranose conformation, which has not been possible in previous structures of polypeptides with C^2-α-D-mannosyl-tryptophan. This conformation is very different to that of alpha-linked *O*-mannosides and has previously been observed in solution by NMR [176]. It can be rationalised by the absence of an anomeric effect in the *C*-glycoside and the steric bulk of the indole moiety making an axial orientation in the usual $^4C_1$ conformation unfavourable. Direct detection of stereochemistry and ring conformation is a major benefit of using an antibody over other techniques like mass spectrometry.

The generation of monoclonal antibodies that recognise C^2-α-D-mannosyl-tryptophan by Western blot will facilitate the study of *C*-mannosylation in various organisms. These antibodies can be used as a means to verify inhibition of *C*-mannosylation activity following the genetic knockout of the *C*-mannosyltransferase gene(s). We have knocked out DPY19L1 and DPY19L3 in THP-1 cells, but these genes can certainly be knocked out in other human cell lines to better understand the implication of this modification in a wide variety of cellular settings (Chapter 7). Knockout mice of DPY19L1 and DPY19L3 have not been published yet and are viable, but that is the next logical step in characterisation of *C*-mannosylation in mammalian model. These antibodies could be used to verify these knockout mice. There are also many opportunities in these reagents in the microbiology field. Since the discovery of *C*-mannosylation in the malaria parasite [124], there has been intense interest in exploring the role of the putative *Plasmodium C*-mannosyltransferase gene. This is also true for *Toxoplasma gondii*, although targeting this gene has proven to be difficult [121]. The key point is that these tools make it very easy to monitor changes in tryptophan mannosylation by Western blot.

Beyond Western blots, the SPR kinetics of antigen binding to each antibody reveals that 6G4, 7G12 and 9E7 may be useful for the affinity purification of mannosylated peptides, as they

have very low k_{off} rates. Given that we have identified bands corresponding to *C*-mannosylated proteins in mammalian serum and Apicomplexan parasite lysate, immunoprecipitations on diverse samples (e.g. human cells, *Toxoplasma gondii* and *Plasmodium falciparum*) seem like a logical next step. For this type of protein modification work, there is always a compromise between antibody affinity and selectivity: ideally one wants high affinity and relaxed enough selectivity to recognise a modification in many contexts. It remains to be seen if these antibodies strike the right balance for this application.

Further refinement of these antibodies for these applications could be achieved through mutagenesis. This could provide more general tools with greater affinity for their ligand. The 5G12 antibody may be the best scaffold to build from because a crystal structure of the antibody-peptide complex is now available and because 5G12 recognises C^2-α-D-mannosyl-tryptophan in the broadest number of peptide contexts. This structure suggests that mutation of $H31_L$ and $S32_L$ in the light chain CDR1 would perturb any interaction with the glutamate in the Glu-Trp(Man) motif. Targeting these positions will likely decrease the antibody's affinity for Glu-Trp(Man) but could broaden or alter its specificity to Xxx-Trp(Man). Site directed mutagenesis at key sites of these antibodies is one way to refine these tools, the other approach could be to generate different peptides using the engineered yeast system and revisit antibody generation to develop complementary tools. The prevalence of the Glu-Trp(Man) motif grants the present tools ample opportunities for discovery: the development of complementary antibody tools is a goal to return to as the field matures.

5.4 - Conclusion

In this chapter, the DPY19 engineered yeast strain from Chapter 2 enabled the multi-milligram production of a peptide bearing C^2-α-D-mannosyl-tryptophan. Conjugation of this peptide to KLH and delivery with Freund's adjuvant provided a high antibody titre in mice, with some animals producing a C^2-α-D-mannosyl-tryptophan-specific response. From these animals, five distinct monoclonal antibodies were obtained, each recognising slightly different glycopeptide sequences and with different binding kinetics. Sequencing data revealed that these antibodies were the result of immunoconvergence: different mice had utilised the same germline alleles to develop a C^2-α-D-mannosyl-tryptophan-specific response. The structure of one of the F_{ab} regions of one of these antibodies in complex with peptide antigen provided insights into how

these antibodies recognise C^2-α-D-mannosyl-tryptophan. These antibodies were finally evaluated on more complex biological samples, including human and bovine serum samples and lysate from an *Apicomplexan* parasite. These experiments demonstrated that the antibodies experienced low promiscuity. As a result, the glycobiology field now has access to high-quality antibody tools for the detection of C^2-α-D-mannosyl-tryptophan and, while these tools have some limitations, their shortcoming are well documented here. These antibodies have been crucial for the development of the first chemical inhibitors of *C*-mannosylation, described in Chapter 6 and are presently being used to identify new endogenously *C*-mannosylated proteins in human and mouse tissues.

Chapter 6

Probing the substrate preferences of DPY19 for the development of novel *C*-mannosyltransferase inhibitors

6.1 - Introduction

The discovery of *C*-mannosylation quickly led to the identification of the WXXW glycosylation consensus sequence (sequon) for this modification [13]. Glycosylation predominantly occurs on the first tryptophan, although it is not uncommon to see the modification on the second tryptophan. In the pursuing years, anomalies were found where *C*-mannosylation was occurring on tryptophan residues that were not part of a WXXW motif. For example, TPOR was discovered to be *C*-mannosylated on Trp^{416} of a $W^{416}AAQ$ motif. Anomalous *C*-mannosylation was also observed in Bovine Lens fibre membrane intrinsic protein (lim2 or MP20) which does not have any WXXW motifs: instead *C*-mannosylation is observed on the $W^{42}RYC$ and $W^{61}NAT$ motifs [109]. These observations appear to suggest that *C*-mannosyltransferases are not absolutely specific to the WXXW motif. How promiscuous these enzymes are, and to what extent they recognise the general amino acids, X in the WXXW sequon, remain unknown. This is a question that could be addressed using my engineered yeast and the *in vitro* assay described in Chapter 2.

An understanding of the substrate preferences of the DPY19 enzymes does not only tell us what proteins are being modified by DPY19, it can also inform the development of substrate-mimicking inhibitors of these enzymes. Inhibitors of *C*-mannosylation will allow one to explore the function of *C*-mannosylation in different cell lines without having to genetically perturb the system. Such a tool could be very useful for defining the role of *C*-mannosylation in diverse systems.

6.2 - Results

6.2.1 - Substrate preferences of the *C. elegans C*-mannosyltransferase

When tryptic peptides of UNC5 containing the $W_1XXW_2XXW_3$ motif were analysed by tandem mass spectrometry in Chapter 2, some intriguing observations were made (Figure 6.1A). DPY19 was able to *C*-mannosylate all of the tryptophan residues in that peptide, but not to the same extent. The peptide was observed in the unmodified state, the singly, doubly and triply modified states. For the singly *C*-mannosylated peptide, instead of modifying the W_1 residue in the $W_1SSW_2SDW_3$ peptide, it exclusively modified the W_2 residue (Figure 6.1A). The only difference between glycosylating the W_2 over the W_1 is that there is an aspartic acid in the X_2 position of the WXXW motif rather than a serine.

This result raises the question of: does DPY19, or *C*-mannosyltransferases more broadly, have a preference for X_1 and X_2 of the WX_1X_2W motif? To test this question, we conducted a BLAST search for all secretory proteins from *C. elegans* and humans with a WX_1X_2W motif, which are putative DPY19 substrates. An amino acid frequency plot was constructed to illustrate that serine was most favored at X_1 followed by threonine for both *C. elegans* and human proteins (Figure 6.1Bi-iii). From this information, we designed and synthesized a series of WXXW peptides that were abundant in the proteome and decided to evaluate the preference of DPY19 for these peptides.

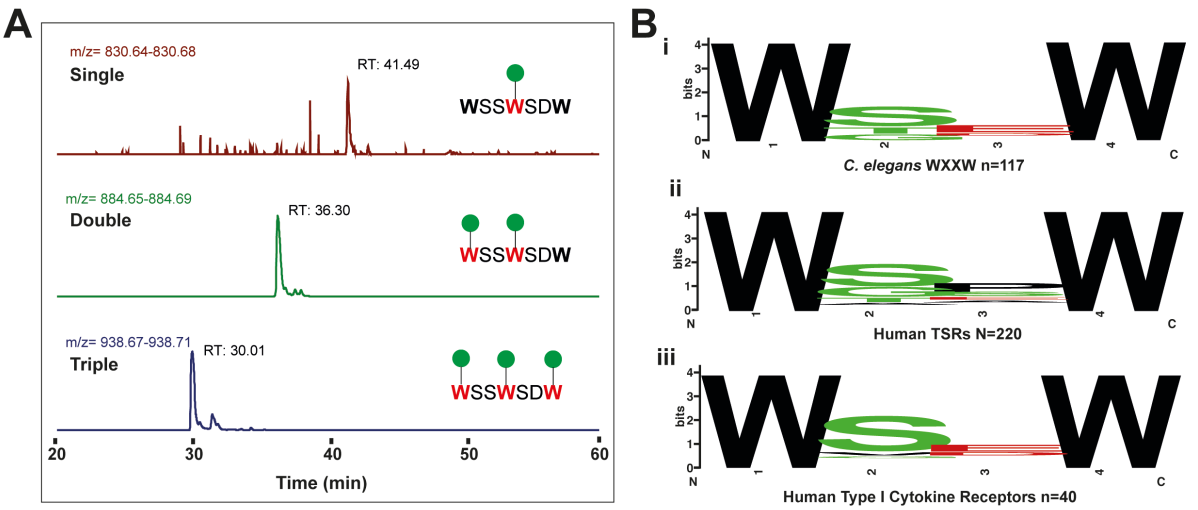


Figure 6. 1: DPY19 may show a preference in amino acids in-between the two tryptophan's

a) Extracted ion chromatogram for the UNC5 glycopeptide data showing different glycoforms.
b) i. Web logo for WXXW motif in *C. elegans*. A total of 117 sequences used.
ii. Web logo for the WXXW motif for human TSR domain containing proteins. A total of 220 amino acid sequences were used.
iii. Web logo for the WXXW motif for human type I cytokine receptors. A total of 40 amino acid sequences were used.

The *in vitro* radioassay developed in Chapter 2 provide a means to test if any of these peptides were preferentially modified by DPY19. A range of WXXW peptide analogues were synthesized, though many were found to have very poor solubility in aqueous buffer (Figure 6.2A). This led to an evaluation of the assay's DMSO tolerance. Figure 6.2B shows the DMSO tolerance of the assay using the WAKW peptide. DMSO concentrations up to 1% are well tolerated without significant impact on enzyme activity. Since our bioinformatics sequence searches showed that WSEW was enriched in both the *C. elegans* proteome and the human type I cytokine receptors, we decided to compare this peptide with WAKW. At a high peptide concentration of 1 mM, DPY19 exhibited a slight preference for WAKW (100%) over WSEW (84%) (Figure 6.2C). At a peptide concentration of 50 μM, there was a significant difference in the preference of WAKW (100%) over WSEW (70%) (Figure 6.2C). This gave us some evidence that just because WSEW is enriched in the proteome, it does not mean that it is preferable for *C*-mannosylation.

We wanted to also investigate the peptide WGGW because in chemical terms, it represents the simplest amino acid structures in between the two tryptophan residues. It would allow us to examine if an amino acid side chain is necessary for recognition. The WGGW peptide was not soluble at 1 mM, but at 50 μM DPY19 was active on the peptide and it had comparable activity to WSEW (70%) but less active as compared to WAKW (Figure 6.2C). We also wanted to investigate if different stereoisomers of tryptophan residues would be recognised by DPY19 which would allow us work out a possible mechanism of how DPY19 engages with the substrate. We wanted to work out the mode of binding to substrate; does it bind to the first tryptophan and then proceeds to the second tryptophan? We synthesized wSEW (first tryptophan is D-tryptophan), WSEw (second tryptophan is D-tryptophan), and wSEw (both tryptophan residues are D-tryptophan). However, the peptides were relatively insoluble at 1

mM and even at 50 μM they were only slightly soluble. Therefore, we could not make any conclusion from our results on these peptides.

Despite the establishment of the WXXW consensus sequence as the minimum requirement for *C*-mannosylation over two decades ago, we wanted to investigate if the second tryptophan residues in the WXXW motif is essential. We synthesized the peptides WSEF and WSEA and examined their activity against WAKW and WSEW. At a high peptide concentration of 1 mM, DPY19 exhibited the same preference for WSEW (84%) and WSEF (89%), with a significantly lower preference for WSEA (72%) (Figure 6.2C). At a peptide concentration of 50 μM, there was a considerably larger difference in the preference for WSEF (79%) over WSEA (46%) (Figure 6.2C). Interestingly, DPY19 can glycosylate WXXZ peptides where Z≠W when the peptide is at high concentration.

We wanted to confirm this result by conducting a cell-based assay where we made two mutants of the WSEW motif in IL12B, WSEF and WSEY (Figure 6.2D). We then transformed the mutants in the DPY19 strain and expressed and purified the IL12B mutants and examine their glycosylation using mass spectrometry (Figure 6.2D). Figure 6.2E shows the extracted ion chromatogram and mass chromatogram for the WSEF mutant, where we see glycosylation of the first tryptophan residue, confirming our *in vitro* C-mannosylation assay result. The WSEY is also *C*-mannosylated, however to a lower extent (Figure 6.2F). The IL12B-WSEF mutant is 71.8% glycosylated, which is significantly higher than the WSEY mutant at only 18.6%. This challenges the long-established notion that there must be a second tryptophan present in the consensus sequence in order for the first tryptophan to be *C*-mannosylated.

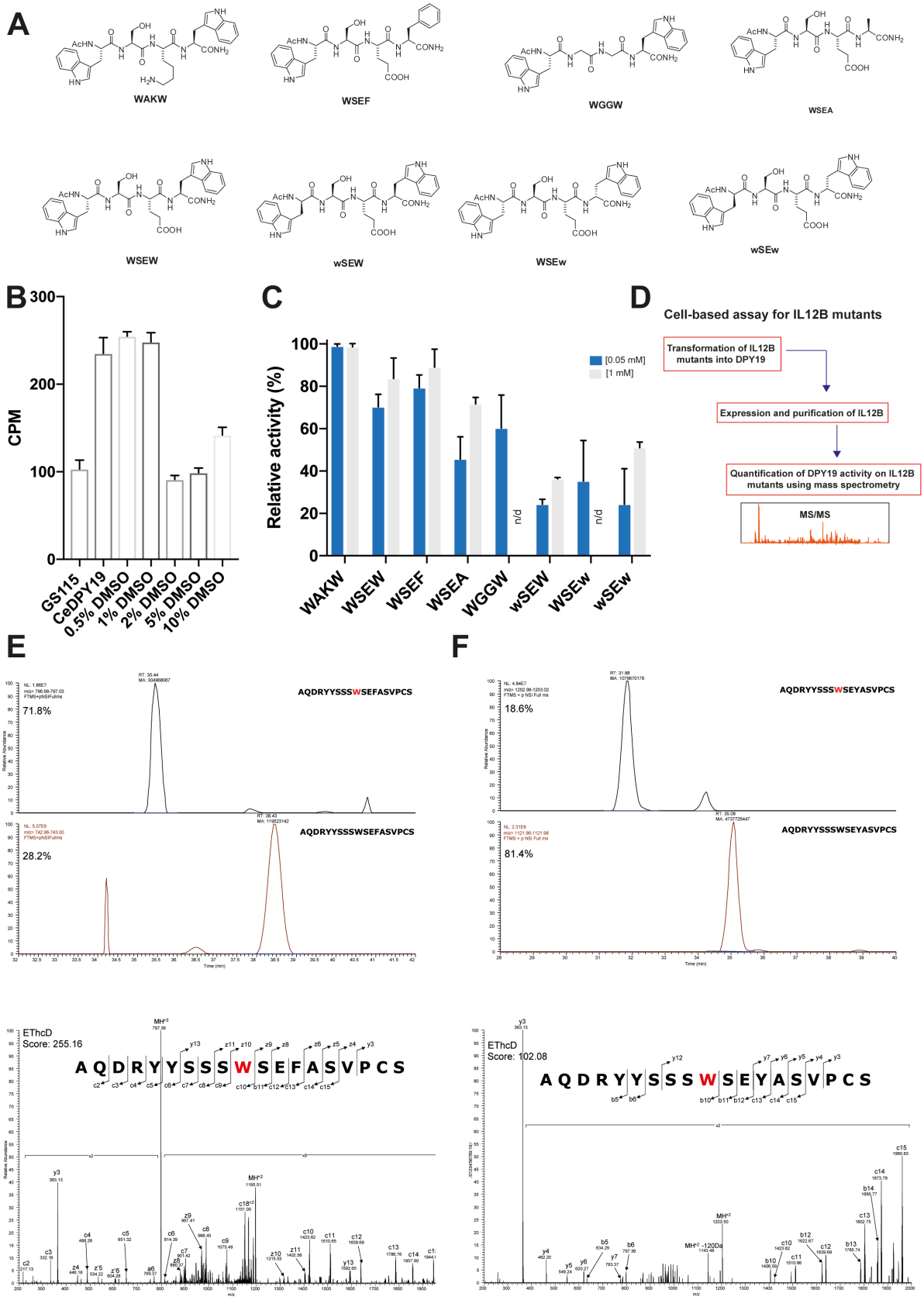


Figure 6. 2: The second tryptophan residue in the WXXW motif is not essential for *C*-mannosylation by DPY19

a) Chemical structures for the different WXXW peptide analogues used in the *in vitro C*-mannosylation assay.
b) Optimisation of the in vitro *C*-mannosylation assay for different concentrations of DMSO.
c) *In vitro C*-mannosylation assay for the different WXXW peptide analogues. All data is in triplicates and the data is normalised relative to the WAKW, where WAKW is given 100% activity and the other peptides are given a value based on their activity to WAKW.
d) Strategy for the cell-based assay confirming *C*-mannosylation of IL12B when the second tryptophan residue in the WSEW motif is mutated to a F or Y.
e) Extracted ion chromatogram and corresponding mass spectra for the AQDRYYSSSWSEFASVPCS peptide for the WSEF mutant.
f) Extracted ion chromatogram and corresponding mass spectra for the AQDRYYSSSWSEYASVPCS peptide for the WSEY mutant.

6.2.2 - Discovery of novel DPY19 peptide-mimetic inhibitors that inhibit *C*-mannosylation in a cell-based assay

The discovery of an inhibitor will allow for the development of tools to probe DPY19 biology and assess the impacts of *C*-mannosylation on biological systems or organisms. Currently, the field of *C*-mannosylation is on a standstill because of the lack of tools to probe the biology of *C*-mannosylation, a chemical inhibitor would greatly advance the understanding of *C*-mannosylation. If it is therapeutically beneficial to inhibit *C*-mannosylation for a particular disease in an organism, then this tool can then be further developed into a pharmaceutical drug in the future. The ability to create different substrate mimetics of the WXXW motif spurred an idea of creating a possible inhibitor of DPY19. Our unique system allows for the production of recombinant proteins with a C^2-α-D-mannosyl-tryptophan. We have also produced antibodies against C^2-α-D-mannosyl-tryptophan, which were introduced in chapter 5 of this thesis. The antibodies along with our engineered system will allow us to test possible inhibitors of DPY19 and *C*-mannosylation.

Generally, when developing inhibitors of a protein target or pathway, there are three ways that are taken to carry out this process. The first one is the high throughput screen where thousands of compounds are assayed for activity against a protein target or cell line. The second is the fragment-based screen, where very small 'fragment' molecules are screened against a protein target and these fragments are 'grown' or 'combined' to yield a higher affinity inhibitor. The last is the rational based drug design approach, where the active site and substrate preferences of a protein target are used to guide the design of compounds that may bind the active site. With our growing knowledge of DPY19 substrate preferences, we reasoned that a rational inhibitor design approach could be fruitful for the development of the first DPY19 inhibitors.

We know what the substrate for *C*-mannosylation is; a WXXW motif where the second tryptophan is important but not necessary. We also know that an α-D-mannose is attached to the C2 carbon of the indole ring of tryptophan. These two facts encouraged the design of substrate mimics that block the C2 carbon of the indole ring of tryptophan or that altered the sterics around or electronics at this position. We reasoned that these molecules would still bind to DPY19 but may not be 'turned over' by the enzyme. We chose four substrate mimics: the benzimidazolone compound (1), the 2-methyl-tryptophan compound (2), the 1-methyl-tryptophan compound (3) and the benzimidazole compound (4) (Figure 6.3A).

To assay these compounds, the yeast strain harbouring DPY19 and a reporter protein such as IL21R was grown for 24 hours, then protein expression was induced with or without the compounds. After 24 hours, the supernatant was harvested and probed for the presence of *C*-mannosylated reporter protein via a western blot or ELISA using the antibodies against C^2-α-D-mannosyl-tryptophan (5G12) (Figure 6.3B). Before testing the substrate mimics, we first tested the substrate to verify that the peptide substrate itself would not act as apparent inhibitor through competition with the reporter protein. The WSSW substrate was tested in the cell based assay and the results indicate that it does not inhibit *C*-mannosylation (Figure 6.3D, E). We then tested the substrate mimic compounds at 1 mM using IL21R as a reporter protein and found that out of the four compounds, only compound 3 inhibited *C*-mannosylation.

It was surprising that compound 3 inhibited DPY19 because, unlike the other compounds, the C2 carbon of the indole ring was not blocked. The implications of this inhibition on the DPY19 enzyme mechanism are unclear. It could be that the steric bulk of the methyl group on the tryptophan nitrogen is sufficient for binding to the enzyme but blocks the subsequent reaction

or that deprotonation or electrophilic attack at the indole nitrogen is an important part of the enzyme's mechanism.

We then wanted to confirm that compound 3 was inhibiting *C*-mannosylation hence we tried mindin as another reporter protein. Mindin has a Flag-tag therefore we can probe with an anti-Flag antibody to confirm if total cellular protein synthesis or translation is not being affected by compound 3 (Figure 6.3H). The results suggest that compound 3 is indeed inhibiting DPY19, and total protein synthesis or translation are likely not being inhibited because we can detect mindin with an anti-Flag antibody but cannot detect presence of *C*-mannosylation using the antibody against C^2-α-D-mannosyl-tryptophan (Figure 6.3I). We do get a significant reduction in detection of mindin with the anti-Flag antibody but we must consider that there is a 4-fold increase in expression of mindin with a C^2-α-D-mannosyl-tryptophan than without the modification as the results indicated in chapter 2 of this thesis. We also tried to examine what the minimum concentration of compound 3 was needed to achieve inhibition. We tried 2 μM and at this concentration there was no significant inhibition (Figure 6.3H, J).

We then wanted to calculate the EC_{50} value for compound 3 because we wanted to design more drug-like and potent inhibitors of DPY19. We titrated the compound from 10 μM to 2000 μM in order to generate a saturation curve. The EC_{50} value for compound 3 was 350 μM (Figure 6.3K). Although this value may be high, it represents a value that works in a cell- based assay and is a starting point for development of more potent inhibitors. This high value can also be attributed to the facts that yeast cells have a thick cell wall and it may be harder for compounds to enter the cell readily (Figure 6.3L). Yeast also have significantly more proteases than mammalian cells therefore the peptide substrate mimic inhibitor might be a target for degradation by peptidases and therefore might need a high concentration to achieve inhibition. It is likely that if compound 3 was tested in mammalian cells, it would achieve a lower EC_{50} value.

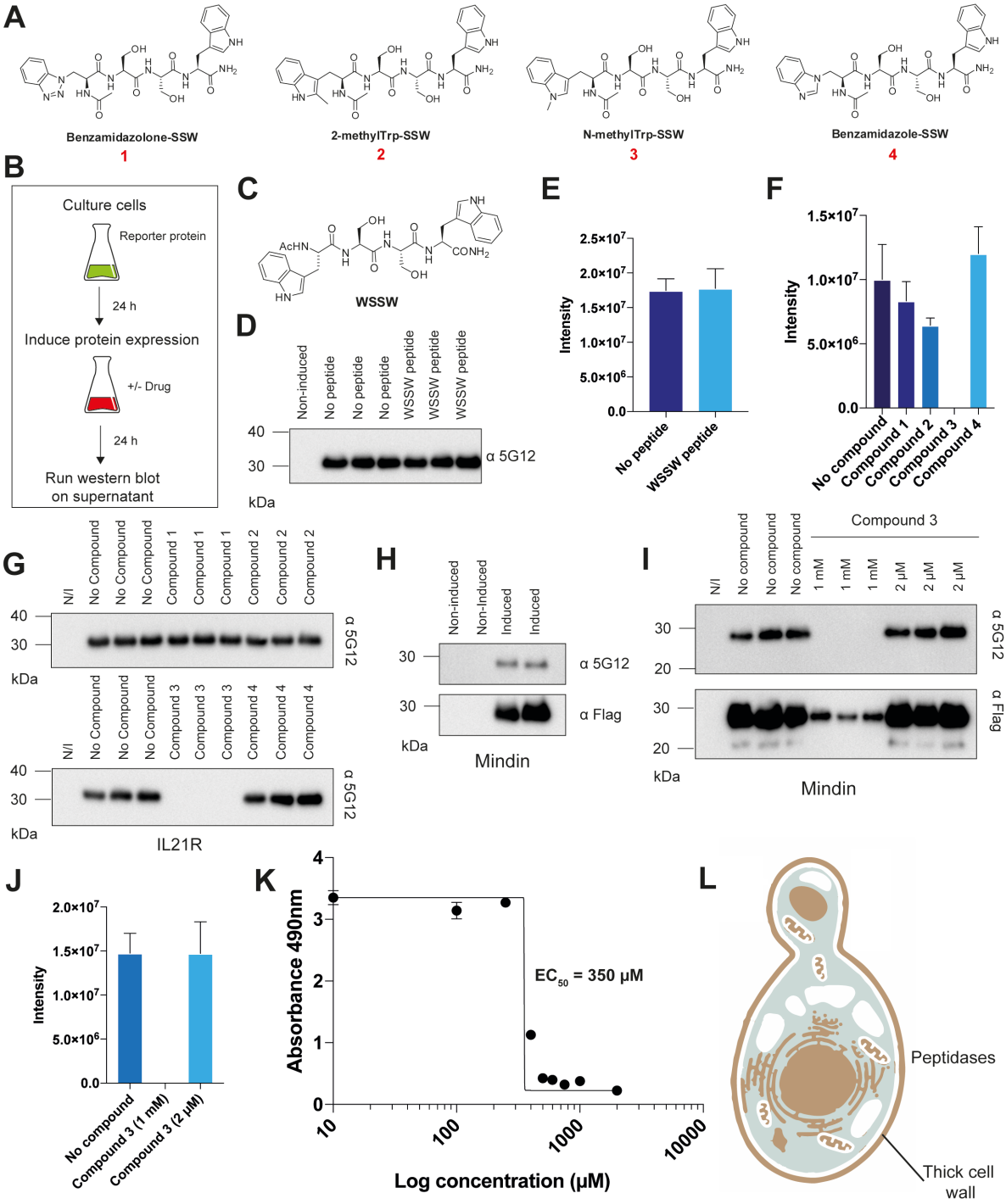


Figure 6. 3: The use of substrate mimics for the development of novel chemical inhibitors of *C*-mannosylation.

a) Chemical structure of the four peptide substrate mimics showing a change in the electronics of the indole ring of tryptophan.

b) Schematics showing how the cell-based inhibitor assay works. Cells harbouring a reporter protein are cultured for 24 hours and then protein expression is induced with or without compounds. After 24 hours supernatant is harvested and an ELISA or western blot is conducted with an antibody against C^2-α-D-mannosyl-tryptophan.

c) Chemical structure of the WSSW peptide substrate
d) Western blot showing the cell-based inhibitor assay results for the WSSW peptide substrate using the 5G12 antibody (against C^2-α-D-mannosyl-tryptophan)
e) Quantification of the cell-based inhibitor assay results for the WSSW peptide substrate in 6.3C
f) Quantification of the cell-based inhibitor assay results for the four peptide substrate mimics in 6.3G
g) Western blot showing the cell-based inhibitor assay results for the four peptide substrate mimics in 6.3A using the 5G12 antibody (against C^2-α-D-mannosyl-tryptophan)
h) Western blot of the cell based assay for mindin showing induced and non-induced mindin using an anti-Flag antibody and 5G12 antibody (against C^2-α-D-mannosyl-tryptophan)
i) Western blot showing the mindin cell-based inhibitor assay results for compound 3 at 1 mM and 2 μM using anti-Flag antibody and the 5G12 antibody (against C^2-α-D-mannosyl-tryptophan)
j) Quantification of the mindin cell-based inhibitor assay results for compound 3 in 6.3I
k) ELISA results quantifying the EC_{50} of compound 3 using the 5G12 antibody (against C^2-α-D-mannosyl-tryptophan)
l) Illustration of a yeast cell highlighting the thick cell wall and enhanced secretion of peptidases.

In order to change compound 3 from a tetrapeptide into a more drug-like molecule, we synthesized six second generation inhibitors of DPY19. We were focused on three regions of compound 3, the first one is the indole ring of the first tryptophan residue, the second consists of two serine amino acids, and the last tryptophan residue (Figure 6.4A). We wanted to explore whether changing the methyl group in the C1 position of the indole ring into other chemical groups would be more beneficial. We also wanted to simplify the sidechains of the two serine amino acids into either glycine (compound 5) or have a 'linker' chain such as the 5-aminovaleric acid (compound 6) (Figure 6.4B). Compound 8 has a benzene group instead of the methyl group (Figure 6.4B). We also wanted to place a methyl group on the C1 position of the second tryptophan residue and examine if that would make the compound more potent (compound 7) (Figure 6.4B). Compound 9 has a 3-indolepropionic acid (IPA) moiety with a methyl group in the C1 carbon position, essentially eliminating the acetyl group. We also

wanted to investigate if there was a need for the second tryptophan residue to be present for recognition (compound 10) (Figure 6.4B).

We then tested the six second generation compounds in our cell based assay and all of the compounds inhibited at 1 mM except for compound 5, with two glycine residues instead of serine residues. It was surprising that compound 6, with the 5-aminovaleric acid, inhibited because it means that the two amino acid residues are not need for recognition (Figure 6.4C). Compound 6 no longer makes the compound into a peptide, but a small molecule. It was not surprising that the second tryptophan molecule was not needed, as compound 10 gave inhibition (Figure 6.4F). Our previous data suggested we do not need the second tryptophan in order for the C-mannosylation to occur. We do not know if these 5 compounds will be more potent than compound 3, as we have not calculated the EC_{50} value for these compounds.

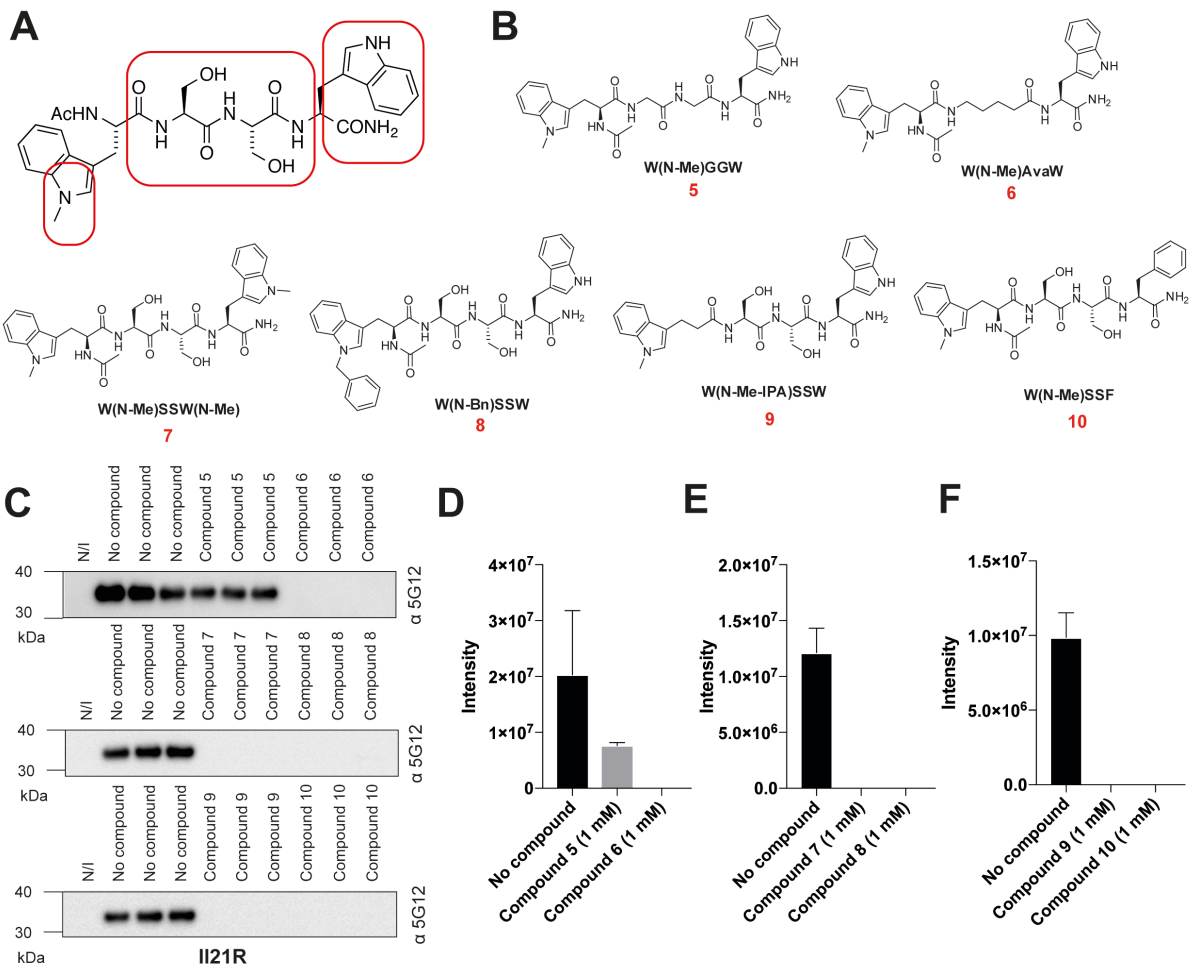


Figure 6. 4: Design of second-generation inhibitors of *C*-mannosylation

a) Schematic illustrating the process of transforming compound 3 from a tetrapeptide to a more drug-like and more potent molecule. The circles represent regions that will be modified.
b) Second generation inhibitors of DPY19 based on compound 3
c) Western blot showing the cell-based inhibitor assay results for the six second generation peptide substrate mimics using the 5G12 antibody (against C^2-α-D-mannosyl-tryptophan)
d) Quantification of the cell-based inhibitor assay results for compound 5 and 6 in 6.4C
e) Quantification of the cell-based inhibitor assay results for compound 7 and 8 in 6.4C
f) Quantification of the cell-based inhibitor assay results for compound 9 and 10 in 6.4C

6.3 - Discussion

Through our susbtrate preference studies, we have found that that there is a preference for X_1 and X_2 in the WX_1X_2W motif. Our bioinformatics analysis found that WSEW was enriched in both the *C. elegans* genome and the human genome but this was not necessarily favourable for *C*-mannosylation. It was likely that those amino acids were playing a structural role for the protein. Our results indicated that WAKW is more favourably glycosylated than WSEW. This means that if we can figure out which amino acids are most favoured and least favoured for the X_1 and X_2 positions in the WX_1X_2W motif, we would be able to predict the level of *C*-mannosylation. Of course, this would not be entirely accurate because protein folding dynamics must be accounted for as well as the level of *C*-mannosyltransferase present but this would allow for prediction of the level of *C*-mannosylation. There have not been any studies shedding any light on this topic, however, there was one study by Hilton *et al* that looked mutating the alanine residue in the WSAW motif of Erythropoietin receptor [59]. The authors mutated the alanine into a glutamic acid residue and found that it had a two-fold increase in cell surface expression. An experiment that we could have tried is comparing *C*-mannosylation of WSEW to WSAW peptides and examining if WSEW *C*-mannosylated to a higher extent than WSAW. If this was in fact that case, then this would explain why this mutant was able to go to the surface more efficiently because it was *C*-mannosylated to a higher extent and was more stable. If it was not, then WSEW might be better for proteins because it plays a structural role.

Another intriguing result discovered in this chapter was that the second tryptophan residue in the WXXW motif is not essential for *C*-mannosylation to occur. For over two decades, it was thought that the second tryptophan was required for *C*-mannosylation. There has only been one account of this anomaly in literature when the thrombopoietin receptor was discovered to harbour a C^2-α-D-mannosyl-tryptophan on a WSEF sequence motif [57]. This result has very wide implications, it opens up the number of proteins with potential *C*-mannosylation. It is very difficult to identify *C*-mannosylation on proteins unless these proteins are purified and are subject to mass spectrometry analysis. This would be a very cumbersome approach if one was trying to identify proteins with *C*-mannosylation that do not contain the non-canonical WXXW motif. An antibody recognising C^2-α-D-mannosyl-tryptophan would be a much more efficient method for identifying these proteins, which we now have.

A question that needs to be answered is what is the minimum requirement for *C*-mannosylation? Is it sufficient to just have exposed tryptophan residues present for *C*-mannosylation to occur? Are there any upstream or downstream motifs present that are recognised by *C*-mannosyltransferases in order for them to modify tryptophan residues? Perhaps other proteins are involved in recognising upstream or downstream motifs and in turn activate *C*-mannosyltransferases to *C*-mannosylate a protein. We could test this hypothesis by taking a protein that is not related to type I cytokine receptor and thrombospondin type I repeat domain containing proteins and place a WXXW motif sequence in that protein. If it gets *C*-mannosylated than there might not be any other motifs for recognition.

Another question to ask is why there are two active *C*-mannosyltransferases in vertebrates and mammals? Both DPY19L1 and DPY19L3 are active and when we align all four DPY19L homologues with the *C. elegans* DPY19 gene DPY19L1 is closer to DPY19 than DPY19L3. If the *C. elegans* DPY19 does not require a second tryptophan in order for it to modify the first tryptophan, then the reason there are two of these active genes in vertebrates and mammals is because they have different substrate preferences. It could be that DPY19L1 evolved to only recognise proteins with the WXXW motif and DPY19L3 evolved to recognise proteins with WXXZ. Evolutionarily, this would be a much more superior method to control which proteins are *C*-mannosylated and which proteins are not. You have two genes that carry out the same function but have selectively different on and off switches. If DPY19L3 is not present in a cell type and there is an abundance of proteins with WXXZ and if it would be a disadvantage if

these proteins were *C*-mannosylated in that cell type than there would not be any negative consequences for that cell type.

The development of inhibitors of *C*-mannosylation will allow the study of *C*-mannosylation to be streamlined. The field of *C*-mannosylation was hampered by a lack of tools, but the addition of an inhibitor of *C*-mannosylation will allow the study of *C*-mannosylation in various cells and organisms. There are many TSR-domain containing proteins in *Plasmodium falciparum* and *Toxoplasma gondii* which have been shown to be *C*-mannosylated [123, 124]. These proteins are mainly secreted proteins involved in invasion and survival; therefore, it may be that a lack of *C*-mannosylation decreases or inhibits their secretion to the cell surface. Our inhibitor may be a new drug lead for treatment of Malaria and Toxoplasmosis because it may impair or kill the parasites. A genome-wide CRISPR screen in *T. gondii* identified that the putative *C*-mannosyltransferase gene was an essential gene and it was lethal when knocked-out, further giving evidence that our inhibitor may be a drug lead for Toxoplasmosis [121] .

6.4 - Conclusions

Our structure activity relationship studies with different WXXW peptide substrates revealed that DPY19 indeed does have different affinities towards different substrates. We uncovered that the second tryptophan residues in the WXXW motif is not essential for *C*-mannosylation to occur. Furthermore, we have discovered the first peptide substrate mimic that inhibits DPY19 in a cell based assay. This chemical tool can now be used to probe the biology of *C*-mannosylation in cell-based systems and organisms.

Chapter 7

Insights into the impact of *dpy19l1* and *dpy19l3* on protein expression

7.1 - Introduction

The first gene to encode a *C*-mannosyltransferase, *dpy19*, was discovered in *C. elegans* in 2013 [68]. The *dpy19* gene family is highly conserved throughout the evolution of metazoans and has undergone substantial diversification [114]. The products of these genes are multi-pass integral membrane proteins and, while some have been functionally validated as *C*-mannosyltransferases, others have no known function. In mammals, there are four *dpy19* homologues (*dpy19l1–4*), which have resulted from gene duplication events throughout evolution. The *dpy19l1* and *dpy19l3* genes resulted from duplication events from an archaic *dpy19 ancestor*, while *dpy19l2* is the result of a more recent duplication of *dpy19l1,* and *dpy19l4* is a more recent duplication of *dpy19l3*. Vertebrates from the reptilia class, only have three of the *dpy19* homologues, which share features with the mammalian *dpy19l1*, *dpy19l3* and *dpy19l4*. Insects, such as *Drosophila melanogaster*, only have one *dpy19* gene, as do the Apicomplexan parasites (*Toxoplasma gondii*, *Plasmodium falciparum* etc.), which likely acquired *dpy19* through horizontal gene transfer. While it is clear that most of these genes encode *C*-mannosyltransferases, it is less obvious what their substrate preferences or biological roles are.

A functional assignment of human *dpy19l1* and *dpy19l3* as *C*-mannosyltransferases has paved the way for an interrogation of the role of *C*-mannosylation in human cells [74]. The Bakker lab expressed the two TSR domains of recombinant mouse UNC5A in *dpy19l1, dpy19l3, dpy19l4* triple knockout CHO cells complemented with a single *dpy19* gene (*dpy19l2* is only expressed in the testes). This allowed them to figure out which *dpy19* genes encode functional *C*-mannosyltransferases. In addition to discovering that DPY19L1 and DPY19L3 are active *C*-mannosyltransferases, they also found that they may have different protein substrate preferences. In the both the TSR1-$W_1STW_2TEW_3SVC$ and TSR2- $W_1SPW_2SKW_3SAC$ tryptic peptides, they found that DPY19L1 had a higher propensity to modify W_1 and W_2, while DPY19L3 had a higher chance of modifying W_3. Although this is an interesting discovery,

these findings are limited to a single overexpressed protein in a heterologous system with an overexpressed enzyme.

As yet, there have not been any studies examining the global effects of *C*-mannosylation on the expression levels of endogenous proteins. What happens to endogenous *C*-mannosylated proteins when the *C*-mannosyltransferase genes are knocked out or inactivated?

In this chapter, the known human *C*-mannosyltransferases *dpy19l1* and *dpy19l3* were disrupted in a human cell line using CRISPR-Cas9 and comparative cell surface proteomics used to interrogate the effects of *C*-mannosylation on cell surface protein expression levels. Many proteins predicted to be *C*-mannosylated had severely reduced cell surface expression levels both the *dpy19l1*$^{-/-}$ and *dpy19l3*$^{-/-}$ cells as compared to the parental cell line. Some selectivity in protein substrates was also noted for DPY19L1 and DPY19L3. This simple series of experiments provides insights into the subtle difference in the biological roles of *dpy19l1* and *dpy19l3*, lays the foundation for more extensive studies in other systems, and generates some interesting hypotheses surrounding possible roles for DPY19L1 and DPY19L3 in health, development and disease.

7.2 - Results

7.2.1 – Disrupting *dpy19l1* and *dpy19l3* in the human monocytic cell line THP-1

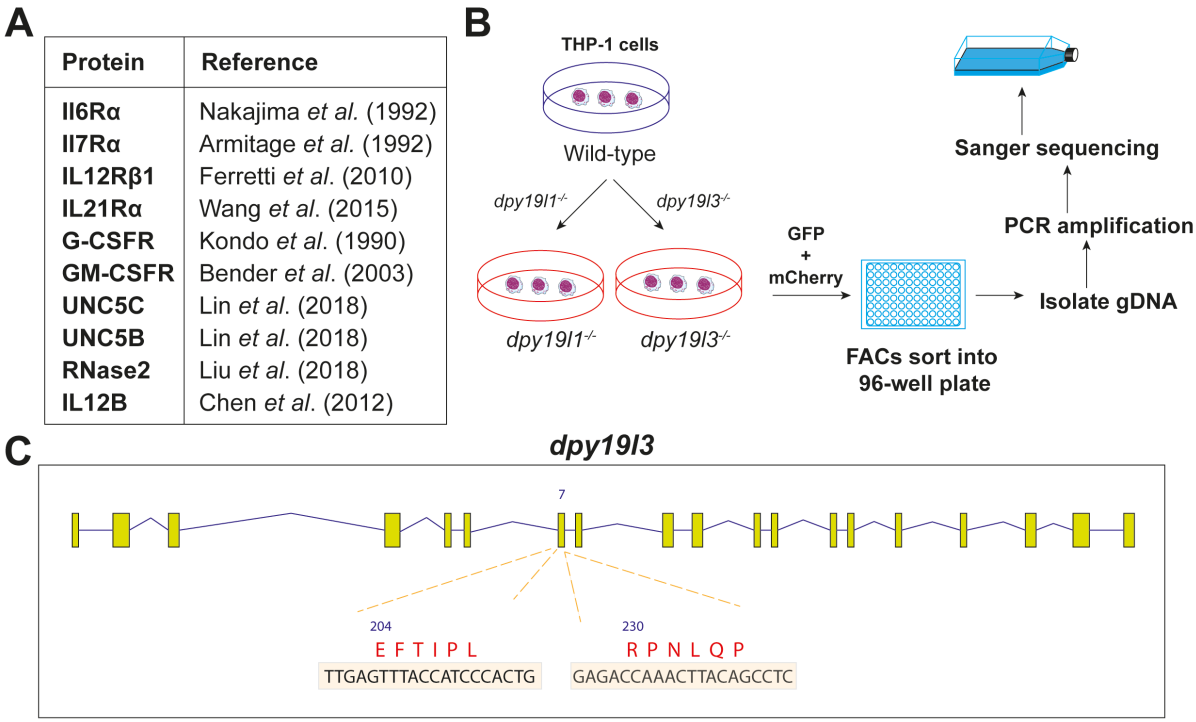
To begin, a literature search was conducted to find an easy-to-culture human cell line that expressed many proteins with the WXXW *C*-mannosylation motif. Evidence was sought at the protein level, not only at the transcript level. THP-1 cells were quickly identified as an ideal human cell line to probe the roles of *dpy19l1* and *dpy19l3*, since they expressed many type I cytokine receptors and thrombospondin type 1 repeats (Figure 7.1A), including: IL6Ra, IL7Ra, IL12R, IL21R, G-CSFR, GM-CSFR, UNC5C, UNC5B, RNaseII and IL12B [177-185].

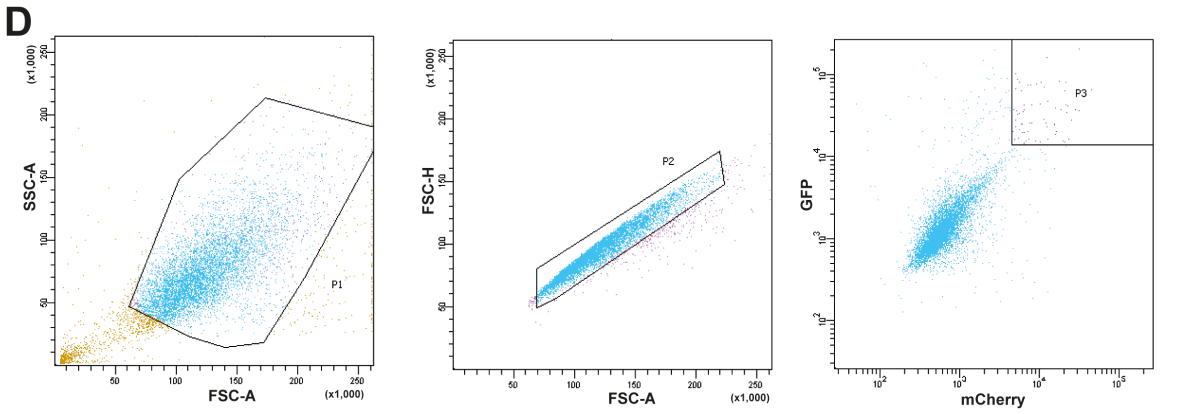
To enable CRISPR-Cas9 disruption of the *dpy19l1* and *dpy19l3* loci, two guide RNAs were designed for each gene. For *dpy19l3*, both guides targeted exon 7 (Figure 7.1C). dsDNA encoding these guides were cloned into a lentiviral vector harbouring a GFP reporter and

transfected into THP-1 cells carrying an inducible Cas9-mCherry gene (Figure 7.1B). Cas9-mCherry expression was induced using doxycycline and flow cytometry used to sort GFP and mCherry positive cells into 96-well plates (Figure 7.1D). These monoclonal cell lines were expanded, their gDNA isolated and Sanger sequencing used to determine if the *dpy19l3* loci had been disrupted at the side targeted by the guide RNA. Eight out of the 10 clones had mutations resulting in truncation of the DPY19L3 protein, five of those clones were heterozygous for the mutation and three clones were homozygous (Figure 7.1E, F). Clone 1, 2, and 7 were used in all subsequent experiments.

A

Protein	Reference
Il6Rα	Nakajima *et al.* (1992)
Il7Rα	Armitage *et al.* (1992)
IL12Rβ1	Ferretti *et al.* (2010)
IL21Rα	Wang *et al.* (2015)
G-CSFR	Kondo *et al.* (1990)
GM-CSFR	Bender *et al.* (2003)
UNC5C	Lin *et al.* (2018)
UNC5B	Lin *et al.* (2018)
RNase2	Liu *et al.* (2018)
IL12B	Chen *et al.* (2012)





E

Clone #	Allele	Protein
Clone 1	Homo null	Truncation
Clone 2	Homo null	Truncation
Clone 3	Het	Truncation
Clone 4	Het	Truncation
Clone 5	No mutation	No effect
Clone 6	Het	Truncation
Clone 7	Homo Null	Truncation
Clone 8	No mutation	No effect
Clone 9	Het	Truncation
Clone 10	Het	Truncation

F

dpy1913$^{-/-}$ Clone 1

```
Wild-type        AAGAGTTGAGTTTACCATCCCACTGAGGGA
Alelle 1 (627)   AAGAGTTGAGTTTACCATC----TGAGGGA
Allele 2 (627)   AAGAGTTGAGTTTACCATC----TGAGGGA
     HOMOZYGOUS NULL - Protein truncation
```

dpy1913$^{-/-}$ Clone 2

```
Wild-type        ACCATCCC..ACTGAGGGAGAACTGGGCGC
Allele 1 (819)   ACCATCCC...CTGAGGGAGAACTGGGCGC
Allele 2 (819)   ACCATCCCACACTGAGGGAGAACTGGGCGC
     HOMOZYGOUS NULL - Protein truncation
```

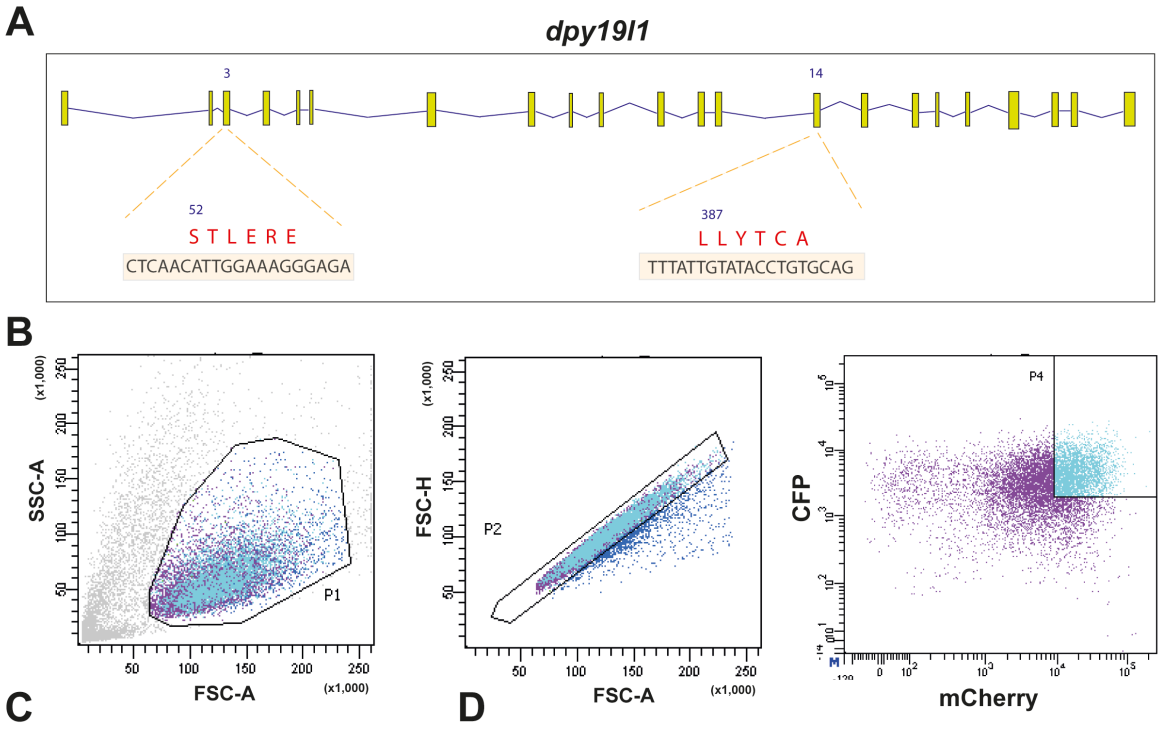
dpy1913$^{-/-}$ Clone 7

```
Wild-type        TACATATTTCCTGAGACCAAACTTACAGTA
Allele 1 (682)   TACATATTTCCTGAG..CAAACTTACAGTA
Allele 2 (695)   TA.........................TA
     HOMOZYGOUS NULL - Protein truncation
```

Figure 7. 1: Establishing *dpy19l3*$^{-/-}$ THP-1 cell lines.

a) Table showing list of proteins with a WXXW motif that are expressed in THP-1 cells with protein level evidence from literature.
b) Diagram showing the strategy of how the CRISPR-Cas9 experiments were conducted.
c) Schematic showing the gene architecture of *dpy19l3* highlighting the introns (lines) and exons (solid bars). The guide RNAs for *dpy19l3* targeted exon 7.
d) Flow cytometry sorting of the doxycycline-induced polyclonal population harbouring guides against *dpy19l3* showing the selection of mCherry and GFP populations.
e) Summary of the sequencing results for 10 of the *dpy19l3*$^{-/-}$ clones, showing Indel generation and status of protein.
f) Alignment of Sanger sequencing for *dpy19l3*$^{-/-}$ clones from above comparing the parental and mutant alleles. Segments highlighted in bold and red are the segments of DNA affects by the Indels.

A similar approach was used to make *dpy19l1*$^{-/-}$ THP-1 cell lines, except the CRISPR-Cas9 lentiviral vector had a CFP reporter protein instead of a GFP reporter, which would facilitate the later production of double-knockouts. For *dpy19l1*, two guide RNAs targeting exon 3 and exon 14 of the were chosen (Figure 7.2A). dsDNA encoding the two guide RNAs were cloned into a CRISPR vector containing the CFP reporter and transfected into THP-1 cells carrying an inducible copy Cas9-mCherry gene. After activation of Cas9 using doxycycline, we sorted single cells into 96-well plates using flow cytometry with CFP and mCherry gated filters (Figure 7.2B). Ten clones were expanded and subjected to Sanger sequencing. Eight of these had mutations resulting in truncation of DPY19L1. Out of the eight clones, three were heterozygous for mutation while five were homozygous null (Figure 7.2C). Clones 1, 3, and 5 for used for subsequent experiments (Figure 7.2D).



C

Clone #	Allele	Protein
Clone 1	Homo null	Truncation
Clone 2	No mutation	No effect
Clone 3	Homo null	Truncation
Clone 4	Het	Truncation
Clone 5	Homo null	Truncation
Clone 6	Het	Truncation
Clone 7	Homo Null	Truncation
Clone 8	Homo null	Truncation
Clone 9	No mutation	No efffect
Clone 10	Het	Truncation

D

dpy1911$^{-/-}$ Clone 1

```
Wild-type        CATTTTTCTCACCTCTCAA-CATTGGAAAG
Allele 1 (425)   CATTTTTCTCACCTCTCA--CATTGGAAAG
Allele 2 (427)   CATTTTTCTCACCTCTCAAACATTGGAAAG
      HOMOZYGOUS NULL - Protein truncation
```

dpy1911$^{-/-}$ Clone 3

```
Wild-type        ATTTTTCTCACCTCTCAACATTGGAAAGGG
Allele 1 (427)   ATTTTTCTCACCTCTCA--ATTGGAAAGGG
Allele 2 (420)   ATTTTTCTCACCTCAC-A--------AGGG
      HOMOZYGOUS NULL - Protein truncation
```

dpy1911$^{-/-}$ Clone 5

```
Wild-type        TCACCTCTCAACATTGGAAAGGGAGATGGC
Allele 1 (433)   TCACCTCTCAT-----GAAAGGGAGATGGC
Allele 2 (433)   TCACCTCTCAT-----GAAAGGGAGATGGC
      HOMOZYGOUS NULL - Protein truncation
```

Figure 7. 2: Establishing *dpy19l1*$^{-/-}$ THP-1 cell lines.

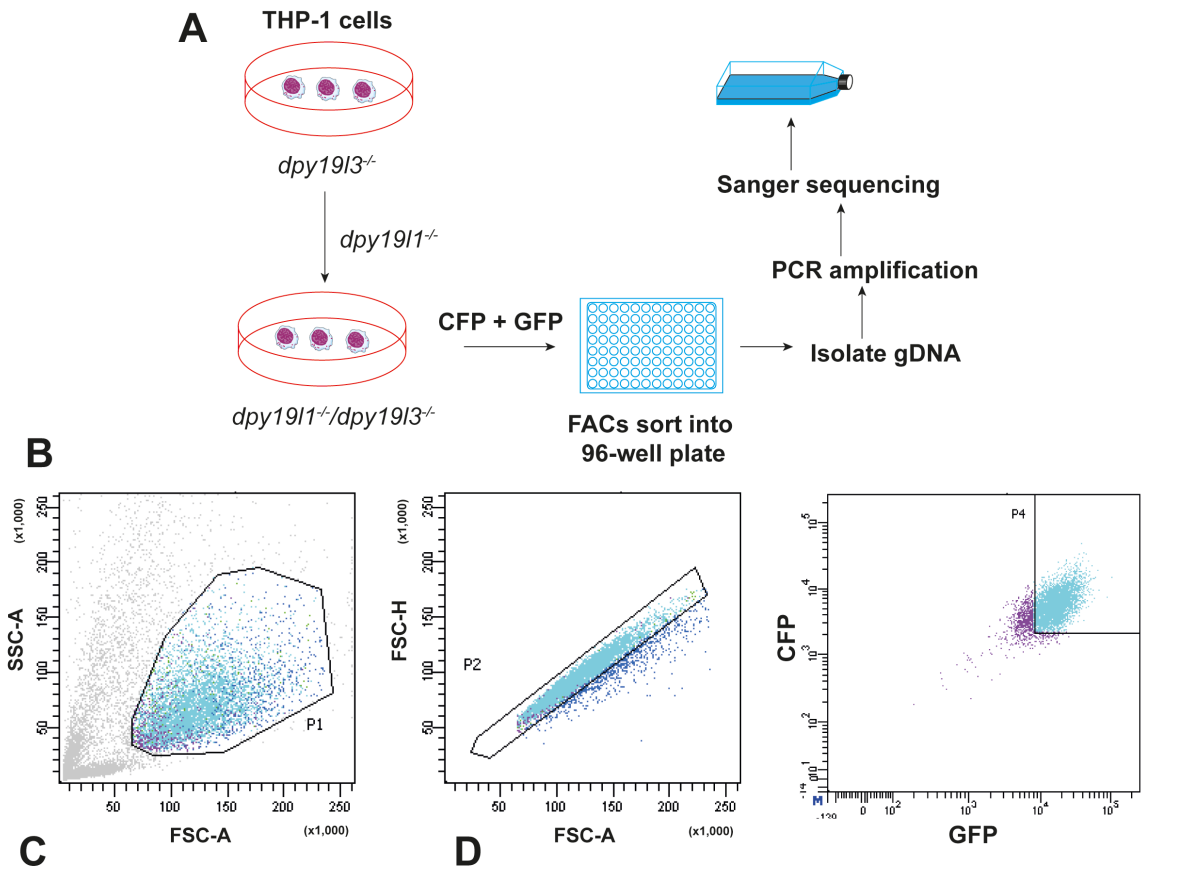
a) Schematic showing the gene architecture of *dpy19l1*$^{-/-}$ highlighting the introns (lines) and exons (solid bars). The CRISPR-Cas9 guide RNAs were designed to target exons 3 and 14.

b) Flow cytometry sorting of the CRSIPR doxycycline-induced polyclonal population showing the selection of mCherry and CFP populations.

c) Summary of the sequencing results for 10 of the *dpy19l1*$^{-/-}$ clones, showing indel generation and status of protein.

d) Alignment of Sanger sequencing for *dpy19l1*$^{-/-}$ clones from above comparing the parental and mutant alleles.

To make the *dpy19l1*$^{-/-}$, *dpy19l3*$^{-/-}$ THP-1 double knockout cell line, *dpy19l3*$^{-/-}$ clone 1 was transfected with the CRISPR cassette targeting exon 3 of *dpy19l1*$^{-/-}$ (Figure 7.3A). Cas9-mCherry was activated using doxycycline and FACS used to clone cells into 96-well plates using the CFP and GFP gated filters (Figure 7.3B). These clones were expanded, genomic DNA isolated for 10 clones, PCR used to amplify the targeted region of *dpy19l1* and this amplicon subjected to Sanger sequencing. Mutations resulting in protein truncation occurred in eight out of the 10 clones. Five of those clones were heterozygous mutations and three of the clones were homozygous null mutations (Figure 7.3C). Clones 7, 8 and 9 were used for subsequent experiments (Figure 7.3D).



C

Clone #	Allele	Protein
Clone 1	Het	Truncation
Clone 2	No mutation	No effect
Clone 3	No mutation	No effect
Clone 4	Het	Truncation
Clone 5	Het	Truncation
Clone 6	Het	Truncation
Clone 7	Homo null	Truncation
Clone 8	Homo null	Truncation
Clone 9	Homo null	Truncation
Clone 10	Het	Truncation

D

dpy1911$^{-/-}$ Clone 7

```
Wild-type        TTCTCACCTCTCAACATTGGAAAGGGAGAT
Allele 1 (430)   TTCTCACCTCTCA---TTGGAAAGGGAGAT
Allele 2 (425)   TTCTC--------ACATTGGAAAGGGAGAT
     HOMOZYGOUS NULL - Protein truncation
```

dpy1911$^{-/-}$ Clone 8

```
Wild-type        ATTTTTCTCACCTCTCAACATTGGAAAGGG
Allele 1 (429)   ATTTTTCTCACCTCTC---ATTGGAAAGGG
Allele 2 (425)   ATTTTTCTCACCTCTC-------GAAAGGG
     HOMOZYGOUS NULL - Protein truncation
```

dpy1911$^{-/-}$ Clone 9

```
Wild-type        GTCATTTTTCTCACCTCTCAACATTGGAAA
Allele 1 (423)   GTCATTTTTCGCACCTCTC----TT---AT
Allele 2 (409)   GTCATTT----------------------AA
     HOMOZYGOUS NULL - Protein truncation
```

Figure 7. 3: Establishing *dpy19l1*$^{-/-}$, *dpy19l3*$^{-/-}$ THP-1 cell lines.

a) Diagram showing the strategy of how the CRISPR-Cas9 experiments will be conducted and how clones for the double knockouts will be generated on the *dpy19l3*$^{-/-}$ clone 1 background.

b) Flow cytometry sorting of the CRISPR doxycycline-induced polyclonal population showing the selection of CFP and GFP populations.

c) Summary the sequencing results for 10 of the *dpy19l1*$^{-/-}$, *dpy19l3*$^{-/-}$ clones, showing indel generation and status of protein.
d) Alignment of Sanger sequencing for *dpy19l1*$^{-/-}$, *dpy19l3*$^{-/-}$ clones from above comparing the parental and mutant alleles.

7.2.2 - Loss of DPY19L1 and DPY19L3 alters cell surface protein expression.

In the earlier chapters of this thesis, some proteins were observed to express better in *P. pastoris* and were more stable when modified by DPY19. Indeed, two to five-fold increases in protein expression levels were not uncommon. Some proteins, like UNC5 and IL21R, failed to express to any detectable level without this modification. As such, it was expected that loss of DPY19L1 and DPY19L3 in human cells would alter the expression profile of secreted and cell-surface proteins with a WXXW motif, which are putative substrates for both of these enzymes. To investigate which proteins were impacted by loss of tryptophan mannosylation, comparative cell surface proteomics was to be conducted on parental, *dpy19l1*$^{-/-}$, *dpy19l3*$^{-/-}$ and *dpy19l1*$^{-/-}$ / *dpy19l3*$^{-/-}$ THP-1 cell lines. The technique utilised here takes advantage of the fact that most cell surface proteins are glycosylated. These *N*- and *O*-glycans usually possess a terminal sialic acid that can be oxidised with sodium periodate to provide a reactive aldehyde functionality (Figure 7.4A). This aldehyde reacts selectively with amino-oxy biotin to covalently tag these glycoproteins with biotin. After biotinylation, cells were lysed and affinity purified using streptavidin immobilised on an agarose resin (Figure 7.4B). After extensive washing steps, these captured proteins are proteolytically digested into peptides, freeing them from the resin, and analysed using LC-MS/MS methods (Figure 7.4B).

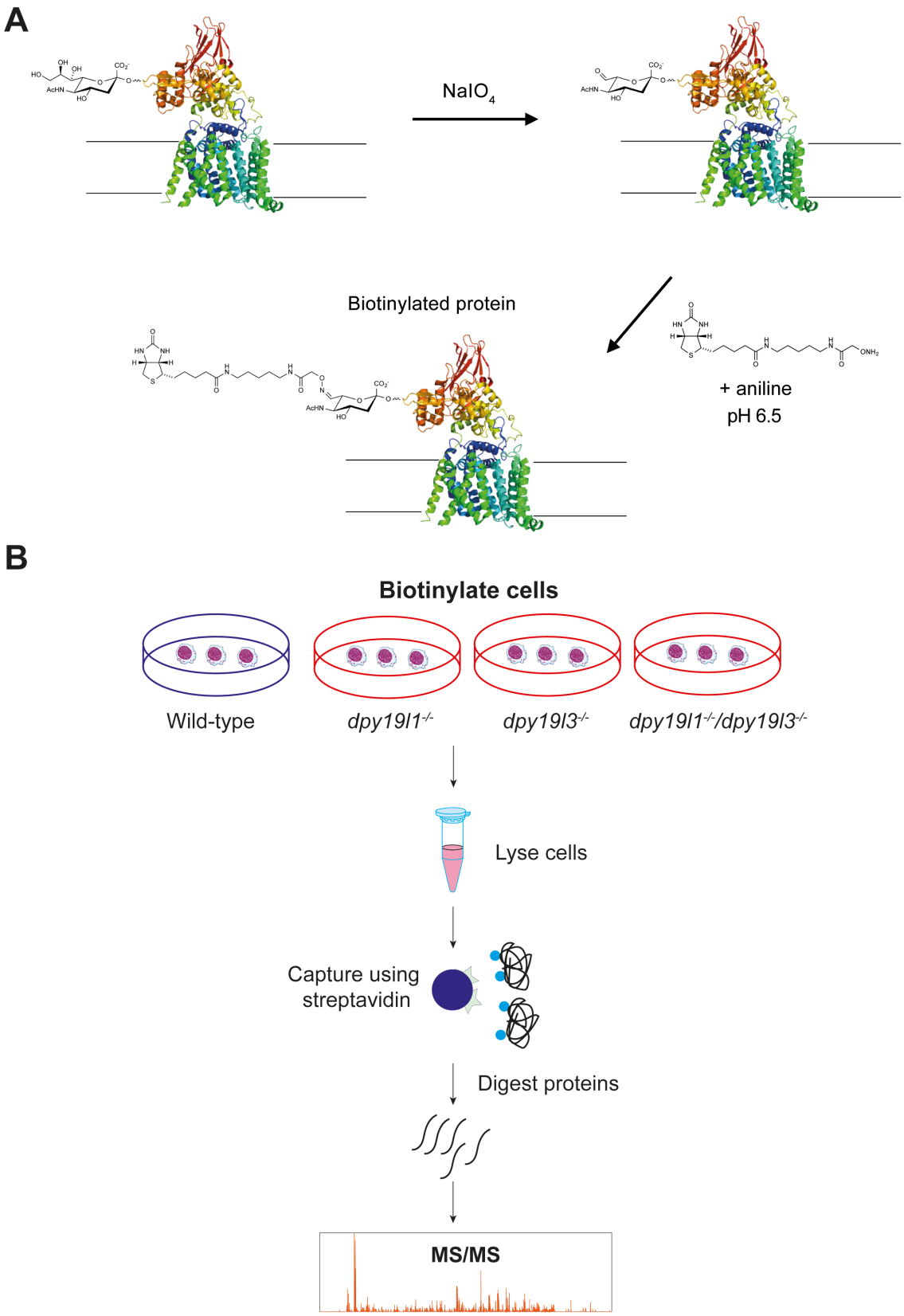


Figure 7. 4: Strategy for biotin-labelling cell surface glycoproteins for affinity purification prior to proteomics.

a) Cartoon showing how the sialic acid on the *N*-glycans of proteins are oxidised with sodium periodate and then labelled with amino-oxy biotin prior to affinity purification.
b) Cartoon showing the work-flow for the proposed cell surface proteomics.

Cell surface proteomics experiments were performed on the THP-1 cells using three independent biological replicates. The samples were prepared by me and the data was collected and analysed by Dr. Jarrod Sandow (WEHI). Figure 7.5A shows a volcano plot comparing results obtained for *dpy19l1*$^{-/-}$ against the parental cell line, where the x-axis represents the $\log_2$ fold change in expression and the y-axis represents the $\log_{10}$ confidence P-value. Proteins represented by blue and red dots are proteins that have a statistically significant change in abundance between the samples. Points in the right quadrant represent those proteins with decreased expression levels in the *dpy19l1*$^{-/-}$ cells, while those in the left quadrant had increased expression levels. Figure 7.5B and C shows the same information for *dpy19l3*$^{-/-}$ and *dpy19l1*$^{-/-}$ / *dpy19l3*$^{-/-}$ cells relative to the parental line, respectively. For all three knockout cell lines, many different proteins have been altered in expression level. However, many of these proteins do not have the WXXW motif, and are therefore not *C*-mannosylated. They most likely have shifted as a result of secondary effects to do with proteins that are *C*-mannosylated. This is also true for the proteins that have increased in cell surface expression. This may be a due to the fact that many proteins with a WXXW motif are cell surface receptor proteins capable of inducing transcriptional changes within the cell (e.g. the type-I cytokine receptors).

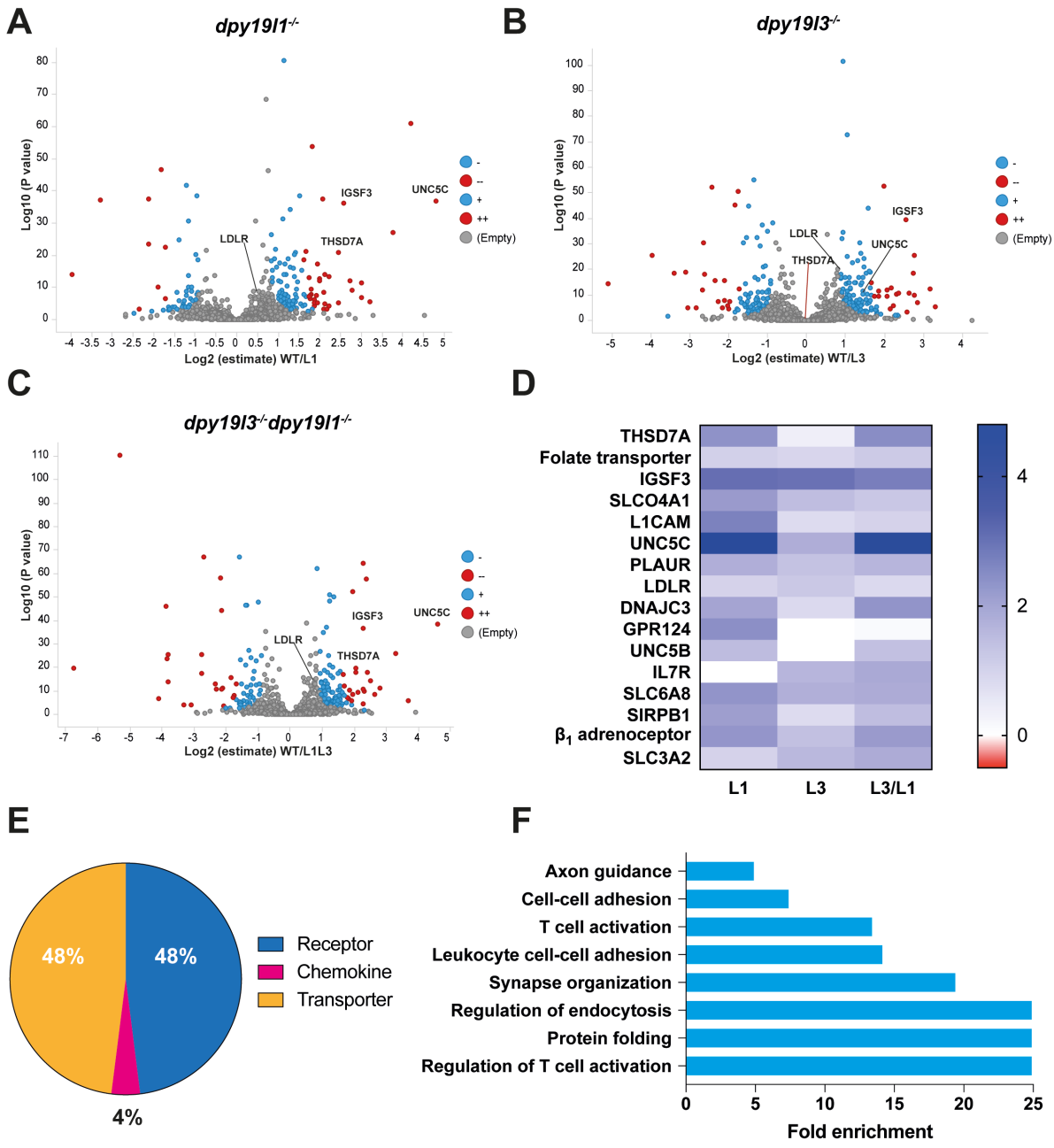


Figure 7. 5: **Comparative cell surface proteomics for cells with disrupted *dpy19l* loci.**

a) Volcano plot comparing the cell surface proteome of parental and *dpy19l1*$^{-/-}$ THP-1 cells. Fold-changes ($\log_2$) of relative protein expression (*x* axis) is plotted against the $\log_{10}$ p-value of the detection (*y* axis). Proteins significantly decreased in expression (right side) or increased in expression (left side) on the THP-1 cell surface are shown in blue (-,+) and red (--,++).

b) Volcano plot comparing the cell surface proteome of parental and *dpy19l3*$^{-/-}$ THP-1 cells. Fold-changes ($\log_2$) of relative protein expression (*x* axis) is plotted against the $\log_{10}$ p-value of the detection (*y* axis). Proteins significantly decreased in expression

(right side) or increased in expression (left side) on the THP-1 cell surface are shown in blue (-,+) and red (--,++).

c) Volcano plot comparing the cell surface proteome of parental and *dpy19l1*$^{-/-}$ / *dpy19l3*$^{-/-}$ THP-1 cells. Fold-changes (log_2) of relative protein expression (*x* axis) is plotted against the log_{10} p-value of the detection (*y* axis). Proteins significantly decreased in expression (right side) or increased in expression (left side) on the THP-1 cell surface are shown in blue (-,+) and red (--,++).

d) Heat map of proteins with a WXXW motif and significant decrease in cell surface expression. L1 = *dpy19l1*$^{-/-}$, L3 = *dpy19l3*$^{-/-}$, L3L1 = *dpy19l1*$^{-/-}$ / *dpy19l3*$^{-/-}$.

e) Pie chart showing the protein classification for the proteins in (d).

f) Gene ontology analysis for proteins with a WXXW motif and significant decrease in cell surface expression in (d).

To begin to make sense of this data, proteins that had a statistically significant decrease in cell surface expression and a WXXW motif were looked at in greater detail. Figure 7.5D shows a heat map for these proteins' fold-change in relative abundance for each of the KO cell lines. This representation reveals some striking differences between DPY19L1 and DPY19L3: many proteins clearly require only DPY19L1 for surface expression, at least one requires DPY19L3 but not DPY19L1, and others require both DPY19L1 and DPY19L3. Some noteworthy proteins include THSD7A and UNC5C. THSD7A, a protein of unknown function, has a ≈4-fold decrease in *dpy19l1*$^{-/-}$ but no significant change in *dpy19l3*$^{-/-}$. Even more striking is the netrin receptor UNC5C, which has a >28-fold decrease in *dpy19l1*$^{-/-}$, while in in *dpy19l3*$^{-/-}$ there is at most a 3-fold reduction in UNC5C surface expression. This striking result suggests that DPY19L1 could play a profound role in netrin signalling.

There are cases where the log_2 fold change in DPY19L1 and DPY19L3 are similar if not the same. IGSF3, another protein of with poorly defined function, had a log_2 fold-change of 2.58 in *dpy19l1*$^{-/-}$, 2.54 in *dpy19l3*$^{-/-}$, and 2.28 in the double knockout, which translates to a 6-fold reduction in protein levels whenever any mannosyltransferase was removed. (Figure 7.5D). PLAUR (Plasminogen Activator, Urokinase Receptor) had a subtler interaction with the mannosyltransferases, with log_2 fold change of 1.44 in *dpy19l1*$^{-/-}$, 1.11 in *dpy19l3*$^{-/-}$, and 1.32 in the double knockout. Also of note was SLC6A8, a creatine transporter, where the log_2 fold

change was 1.92 in *dpy19l1*$^{-/-}$, 1.42 in *dpy19l3*$^{-/-}$, and 1.5 in the double knockout cells (Figure 7.5D).

The best example of a protein being dependent on DPY19L3 but not DPY19L1 was IL7R (the interleukin 7 receptor), which was unchanged in *dpy19l1*$^{-/-}$ and decreased by 2–3-fold in *dpy19l3*$^{-/-}$ and the double knockout cell lines.

All of the proteins that had reduced cell surface expression and a WXXW motif were analysed using functional pathway gene ontology software called PANTHER [26]. Figure 7.5E illustrates the breakdown of the proteins into their respective functional classifications. Surprisingly, there are as many transporters as there are receptors. No integral membrane transporters have yet been shown to be *C*-mannosylated, perhaps because no one has focussed on these proteins, which are harder to work with than the membrane proteins with a single transmembrane region. Some transporters that warrant further investigation for *C*-mannosylation include the folate transporter, the solute carrier organic anion transporter family member 4A1 (SLCO4A1) and sodium- and chloride-dependent creatine transporter 1 (SLC6A8).

The functional pathways of all of these putative mannosyltransferase substrates were noted to interact with a diverse range of unrelated signalling pathways (Figure 7.5F). This likely explains why so many proteins without a WXXW motif change in abundance in the absence of DPY19L1 and DPY19L3.

To expand the list of human proteins that are differentially regulated by *C*-mannosyltransferases, the THP-1 monocytes were differentiated into macrophage-like cells. These macrophage-like cells express and secret different proteins to the monocytes [186]. An established method to differentiate THP-1 monocytic cells into macrophage-like cells makes use phorbol 12-myristate-13-acetate (PMA) (Figure 7.6A) [187]. Using a very high dose of PMA (>100 ng/ml) can induce enhanced secretion of TNF, a pro-inflammatory cytokine, making these cells more sensitive to pro-inflammatory stimuli [188]. Therefore, it is important to use lower concentrations of PMA that will not activate the macrophage-like cells. In this case, differentiation was accomplished using 10 ng/ml of PMA for 48 h, followed by a 24 h rest period without PMA, tryptic detachment of cells and subculture (Figure 7.6B) [188]. Due to time constraints, cell surface proteomics were only conducted on the double knockout and

parental cell lines. To preserve the surface proteins of these adherent cells (undifferentiated THP-1 cells are non-adherent) we used EDTA to detach cells rather than trypsin.

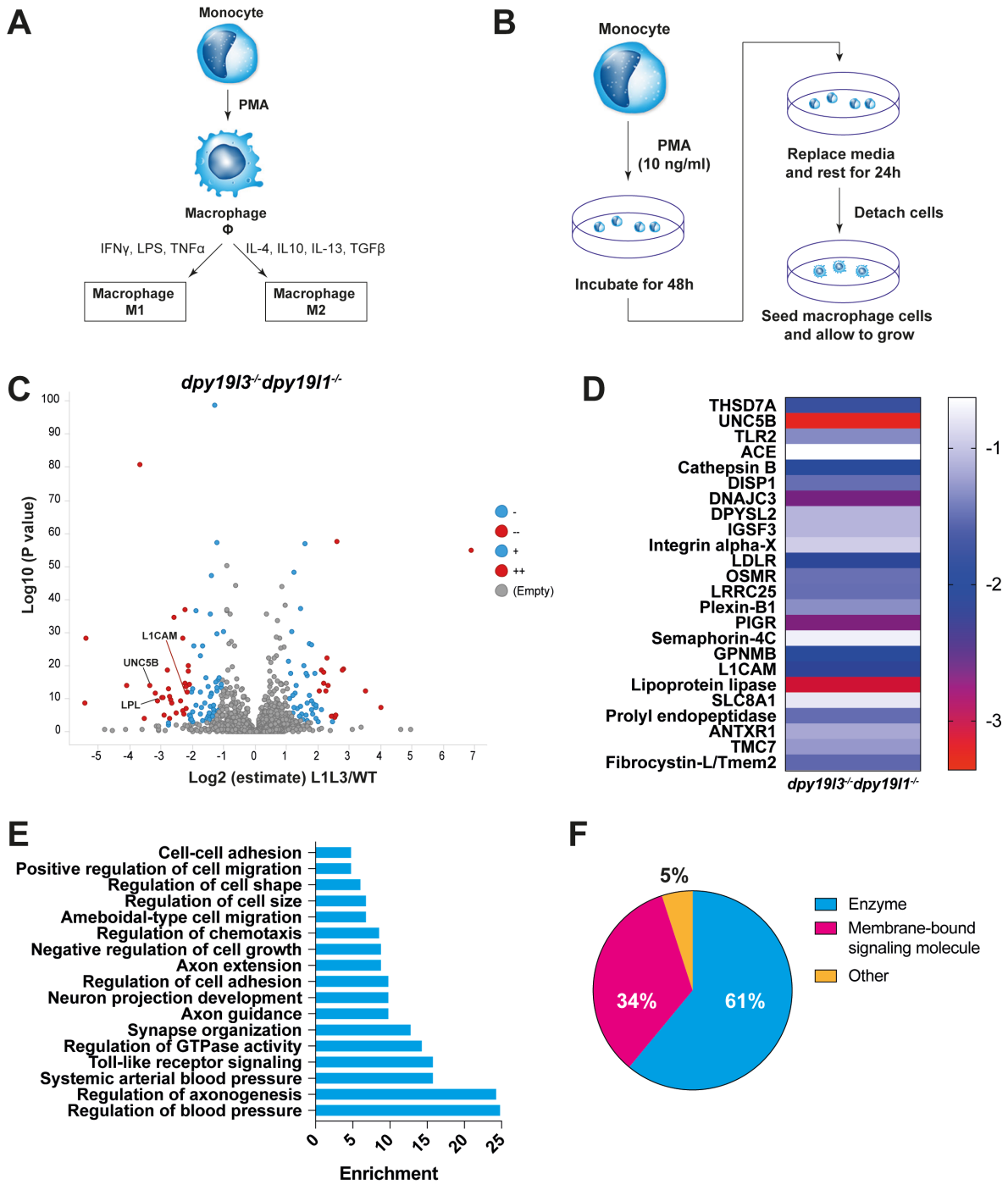


Figure 7. 6: Comparative cell surface proteomics for double knockout and parental THP-1 macrophage-like cells.

a) Work-flow for differentiation of THP-1 monocytic cells into macrophage-like cells using PMA. M1 and M2 macrophages can also be obtained from these cells.

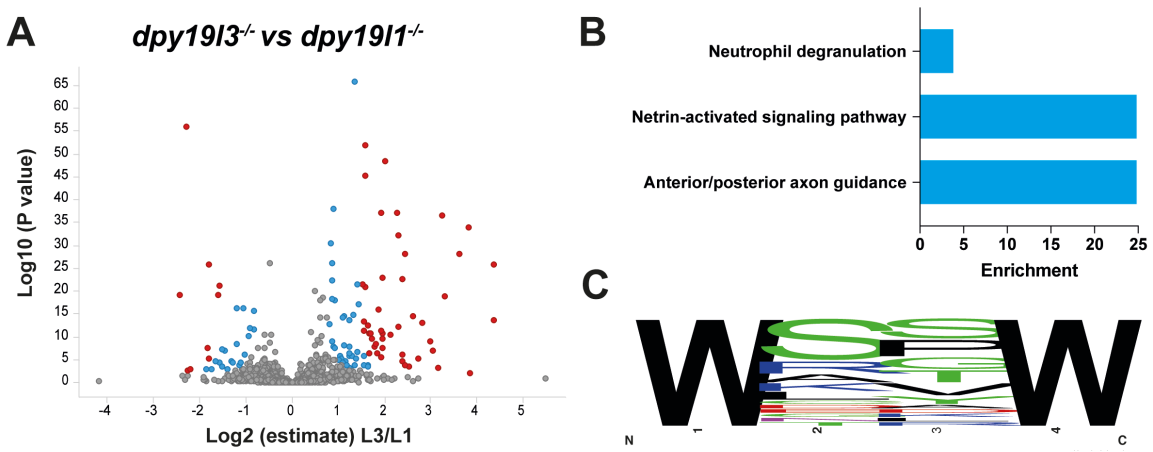
b) Diagram illustrating the strategy for differentiating the parental and double knockout THP-1 cells into macrophage-like cells.
c) Volcano plot comparing the cell surface proteome of parental and double knockout THP-1-derived macrophage-like cells. Fold-changes (log_2) of relative protein expression (*x* axis) is plotted against the log_{10} p-value of the detection (*y* axis). Proteins significantly decreased in expression (right side) or increased in expression (left side) on the THP-1 cell surface are shown in blue (-,+) and red (--,++).
d) Heat map of proteins with a WXXW motif and a significant decrease in cell surface expression in the double knockout THP-1-derived macrophage-like cells.
e) Gene ontology analysis for proteins with the WXXW motif and a significant decrease in cell surface expression (d).
f) Pie chart illustrating the protein classification for the proteins in (d).

In the volcano plot in Figure 7.6C, the proteins with a negative log_2 fold change (the left quadrant) are those that a decreased in the double knockout. The heat map in Figure 7.6D shows the proteins with a WXXW motif that have decreased in cell surface expression level. There are many proteins that have a decreased cell surface expression in these differentiated macrophage-like cells that were also observed in the monocytic form (Figure 7.6D). These include: IGSF3, THSD7A, UNC5B, DNAJC3, LDLR, L1CAM, and SLC6A8 (Figure 7.6D). This result provides improved confidence in the significance of this data. Notably, there is no expression of UNC5C in these macrophage-like cells, which was the most severely impacted protein in the monocytic form. Conversely, there are many proteins observed in the macrophage-like cells but not the monocytic cells that possess a WXXW and experience a decrease in cell surface expression. These include: lipoprotein lipase (-3.10, ≈8-fold decrease), polymeric immunoglobulin receptor (PIGR) (-2.70, ≈6-fold decrease), cathepsin B (-1.98, ≈4-fold decrease), transmembrane glycoprotein NMB (GPNMB) (-1.97, ≈4-fold decrease), fibrocystin-L (-1.41, ≈2.5-fold decrease), prolyl endopeptidase (-1.38, ≈2.5-fold decrease), leucine-rich repeat-containing protein 25 (LRRC25) (-1.37, ≈2.5-fold decrease), protein dispatched homolog 1 (DISP1) (-1.37, ≈2.5-fold decrease), oncostatin-M specific receptor (OSMR) (-1.36, ≈2.5-fold decrease), Toll-like receptor 2 (TLR2) (-1.23, ≈2-fold decrease) and Plexin-B1 (-1.21, ≈2-fold decrease) (Figure 7.6D). The majority of these proteins are enzymes (61%) or membrane-bound signalling receptors (34%) (Figure 7.6F). This is quite different to that observed for the monocytic cells, where most of the proteins were receptors

and transporters. These proteins played roles in myriad pathways that were, once again, quite different to that observed in the monocytic cells (Figure 7.6E).

7.2.3 - DPY19L1 and DPY19L3 impact the expression profiles of different proteins

Having observed some differences in the way that DPY19L1 and DPY19L3 impact protein expression, the *dpy19l1*$^{-/-}$ and *dpy19l3*$^{-/-}$ data sets were compared directly to provide a more systematic assessment of the differences between these enzymes. Figure 7.7A is a volcano plot illustrating this information: proteins in the right quadrant (positive $\log_2$ fold-change) are less abundant in the *dpy19l1*$^{-/-}$ cells and proteins in the left quadrant (negative $\log_2$ fold-change) are less abundant in *dpy19l3*$^{-/-}$ cells. This clearly shows that a greater number of proteins are impacted by DPY19L1 than DPY19L3 (Figure 7.7A). This raises the question as to whether DPY19L1 is more active in cells, simply has a greater number of substrates than DPY19L3, or plays a more important role in stabilising proteins.



D

Selecivity	Uniprot ID	Gene name	WXXW motif	Log_{10} P-value
DPY19L1	Q9UPZ6	THSD7A	WSSW,WSTW,WSAW,WSPW	22.7
	Q7Z403	TMC6	WLPW	11.4
	Q8IZJ1	UNC5B	WSSW	5.7
	O95185	UNC5C	WSTW,WTPW	28.3
	P12821	ACE	WRSW,WEGW	28.1
	Q8IZF4	ADGRG5	WALW,WGGW	3.3
	Q96PE1	ADGRA2	WLVW	10.8
	Q92485	ASM3B	WKPW	3.5
	Q13217	DNAJC3	WNSW	1.8
	P15586	GNS	WSYW	2.1
	Q8N6C5	IGSF1	WKKW	21.0
	P20702	ITGAX	WRRW	32.1
	O43157	PLXNB1	WCVW,WAGW	6.6
	O60779	SLC19A2	WSTW,WSVW	3.7
	O15389	SIGLEC5	WRSW	13.8
	Q92673	SORL1	WDDW	5.5
DPY19L3	Q9UKU6	TRHDE	WEPW	4.4
	O00763	ACACB	WAGW	15.5
	Q9HBG7	LY9	WCIW	2.8
	P16871	IL7R	WSEW	2.8

Figure 7. 7: A direct comparison of *dpy19l1*$^{-/-}$ and *dpy19l3*$^{-/-}$ THP-1 cells.

a) Volcano plot comparing cell surface proteomic data for *dpy19l1*$^{-/-}$ and *dpy19l3*$^{-/-}$ cells. Fold-changes (log_2) of relative protein expression (*x* axis) is plotted against the log_{10} p-value of the detection (*y* axis). Proteins that have significantly changed in abundance are shown in blue (-,+) and red (--,++).

b) Gene ontology analysis for proteins with a WXXW motif that show a selective and significant decrease in cell surface expression in *dpy19l1*$^{-/-}$ cells.

c) Web logo for the 16 proteins identified to have a selective decrease in cells surface expression in the *dpy19l1*$^{-/-}$ cells.

d) Table showing the identity of the proteins that are selectively perturbed by loss of DPY19L1 and DPY19L3 and their WXXW motifs.

A table of proteins that are impacted only by DPY19L1 or DPY19L3 is provided in Figure 7.7D, along with the WXXW motif from this protein. DPY19L1 impacts the expression of four times more proteins in monocytes than DPY19L3. The expression of THSD7A, UNC5C, UNC5B, ACE, and DNAJC3 are most dramatically impacted by DPY19L1. For the DPY19L3, the most significant shift was for IL7R. A web logo was constructed for the WXXW sequences for proteins impacted by DPY19L1 (Figure 7.7C), though no significant consensus was observed that might explain the specificity of this enzyme for these proteins. There are reports that DPY19L3 actually recognises the WXXC motif, which may explain this observation [74]. It could also be that the stability of these proteins is differentially impacted by mannose on the first or last tryptophan of the WXXW motif and that the observed dependencies are a function of the different protein folds than the local sequence of the mannosylation consensus motif.

The data presented in this chapter reveals that a loss of either DPY19L1 and DPY19L3 significantly reduces the cell surface expression of many putatively *C*-mannosylated proteins. Whilst this proteomic data is intriguing and has facilitated the development of several hypotheses, these results must yet be validated using orthogonal techniques, such as FACS and western blots and microscopy. Attempts were made to do this for the most down-regulated proteins, UNC5B and UNC5C, but the commercial antibodies we obtained for these proteins failed to recognise anything. With suitable antibodies, microscopy would also be a useful technique for probing these putative protein trafficking defects. Quantitative PCR or RNAseq data would also be useful to show that the reduction in cell surface expression is not due to changes in transcription

7.3 - Discussion

C-mannosylation is proving to be a modification that is crucial for protein folding, stability and/or trafficking. In this chapter, disruption of *C*-mannosylation in human monocytic THP-1 cells severely impeded the ability of many proteins to reach the cell surface. A similar result was observed in macrophage-like cells derived from these THP-1 cell lines. These results mirrored results in chapters 2 and 3 of this thesis, where expression of DPY19 in *Pichia pastoris* significantly improved the expression of proteins with a WXXW motif and proteins with C^2-α-D-mannosyl-tryptophan were markedly more stable than proteins without.

In terms of evolution, it is very interesting to consider why there are at least two functional *C*-mannosyltransferases in mammals and vertebrates. With the exception of *dpy19l2*, all other *dpy19* genes are expressed in all mammalian tissues, suggesting that there is not much variation in their tissue expression profiles. Generally, organisms which have one *dpy19* gene tend to only have TSR domain containing proteins as substrates of *C*-mannosylation. While those with two or more *C*-mannosyltransferases have other *C*-mannosylated proteins, like the type-I cytokine receptors. Our data reveals that most of the proteins dependent on DPY19L1 expression possess a TSR domain, while IL7R, the protein dependent on DPY19L3 expression, is a type-I cytokine receptor. Thus, it may be that the *C*-mannosyltransferases have diversified as their repertoire of substrates has expanded over time. Another complementary explanation for the diversification of *dpy19* genes is that differential expression of DPY19L1 and DPY19L3 provides an alternative means to control client protein expression levels.

One interesting observation from this work in monocytic cells is that there were four times as many proteins dependent on DPY19L1 and there were for DPY19L3. It is hard to interpret the significance of this result, as we have no measure of the absolute number of proteins modified by each enzyme. Could it be that there is a greater number of protein substrates for DPY19L1 as compared to DPY19L3? To what degree are these enzyme's redundant? What are the true consensus motifs for each of these enzymes? These are all important questions that remain to be explored. It may be that DPY19L3 has evolved specifically to glycosylate non-canonical motifs, such as WXXC, and that it plays a more specialised role than DPY19L1 in glycosylating important subsets of proteins. In chapter 6 of this thesis, it was shown that the *C. elegans* DPY19 can *C*-mannosylate non-canonical WXXF/Y motifs, so there is precedence for this non-canonical activity. Shcherbakova *et al.* have reported that DPY19L3 specifically C-mannosylated the W_3 motif in the $W_1XXW_2XXW_3XXC$ sequence of UNC5A in CHO cells [74]. Loss of DPY19L3 and mannosylation of the WXXC motif did not impact UNC5A secretion. By contrast, loss of DPY19L1, which *C*-mannosylates W_1 and W_2, resulted in severe reduction of UNC5A secretion. Mannosylation of the first tryptophans in these proteins appears to be the most important for their stability. This may explain why the stability of more proteins are impacted by the loss of DPY19L1 than DPY19L3. Very little is known about the role of *C*-mannosylation in protein kinesis, which covers events from protein transcription to protein secretion [189]. *C*-mannosylation stabilises proteins but what happens to these proteins in the absence of tryptophan mannosylation? Do they fail to fold in the ER and/or denature along the

protein trafficking pathway or at the cell surface? Are there chaperones that recognise C^2-α-D-mannosyl-tryptophan in a similar vain to calnexin and calreticulin in the ER [190-192]? Unfortunately, we cannot answer this question at this moment, but a good starting point would be to see where these proteins are ending up using a combination of immunofluorescence microscopy and molecular probes to inhibit the proteasome and ERAD.

UNC5C is the protein that was most severely impacted by the loss of *C*-mannosylation by DPY19L1. UNC5C, along with UNC5A, B and D, has two TSR domains and is a receptor for netrin-1, which is involved in apoptosis, cell migration and axon guidance. Rare activating mutations in UNC5C results in increased neuronal cell death and increased sensitivity to cytotoxic molecules such as amyloid-beta [162]. UNC5C mutations are a genetic predictor for the late onset of Alzheimer's disease. It is tempting to speculate that inhibition of DPY19L1 may restrict trafficking of this death receptor to the cell surface and slow neuronal cell death in these patients. Further exploration of this concept in neuronal cell lines could yield new directions for the treatment of a subset of people with familial Alzheimer's disease.

Most of the other proteins that had significant reduction in cell surface expression are relative unknowns. Thrombospondin type I domain-containing protein 7A (THSD7A) is a cell surface protein that is involved in endothelial cell migration and cytoskeletal organisation [193]. When THSD7A was knocked out of oesophageal squamous cell carcinoma (ESCC) cells lines, reduced tumour cell invasion, migration and proliferation was observed [194]. Thus, it may be that inhibition of DPY19L1 could minimise the metastasis of some tumours. Immunoglobulin superfamily, member 3 (IGSF3) is another protein for which little is known. It is a transmembrane protein and is involved in neuronal morphogenesis [195]. Lastly, DnaJ homolog subfamily C member 3 (DNAJC3) is a chaperone protein involved in stress responses and an absence of this protein results in diabetes mellitus and multi-systemic neurodegeneration [196, 197], suggesting that inhibition of DPY19L1 could have deleterious side-effects.

Lipoprotein lipase (LPL) is another interesting protein that had a significant reduction in cell surface expression in the DPY19L1$^{-/-}$ DPY19L3$^{-/-}$ double knockout macrophage-like cells. This protein had an 8-fold reduction in cell surface expression and it was recently discovered that this protein was *C*-mannosylated [97]. The authors of the paper mutated tryptophan residues of the WXXW motif and claimed that the loss of *C*-mannosylation of this protein causes

decreased secretion and enzymatic activity. However, mutating tryptophan residues in the WXXW motif could have confounding results because the mutations themselves can destabilise the protein or cause it to not fold properly. Therefore, it cannot be claimed that *C*-mannosylation is essential on this data alone. The result in this chapter provide unambiguous evidence that a loss of *C*-mannosylation reduced LPL cell surface expression. Whether *C*-mannosylation impacts the enzymatic activity of this protein remains to be seen.

Another interesting protein with reduced cell surface expression is the Low-density lipoprotein receptor (LDLR), which cluster in clathrin-coated pits. The main role of this protein is to function as a receptor for Low-density lipoprotein (LDL), which is an important physiological process. This protein also acts as an entry point for the hepatitis C virus and other viruses into the liver [198],[199]. This is an important drug target for managing LDL levels and virus infections: DPY19 inhibition may bring a new dimension to LDLR drug discovery initiatives.

There were various other proteins that had significant reduction in cell surface expression and both of these proteins have a role in cancer progression. L1CAM had both a reduction in the monocytes and macrophage-like cells. L1CAM is involved in cell migration and cell adhesion and is implicated in cancer metastasis, in particular melanoma invasion [200]. The reduction of cell surface expression of this protein will also reduce tumour metastasis. IL7R is the other interesting protein involved in cancer which had a decrease in cell surface expression when DPY19L3 was knocked out. Activating mutations in IL7R drive proliferation of T-cell acute lymphoblastic leukaemia (T-ALL) [160]. Inhibition of DPY19L3 could reduce the cell surface expression of this mutant IL7R and may slow the progression of T-ALL.

To accompany these musings on the possible therapeutic roles of DPY19 inhibitors inferred from what's known about the proteins identified by these proteomic experiments, it is worth considering the similarities and differences of DPY19L1 and DPY19L3. These enzymes share 29% similarity at the amino acid sequence level but clearly recognise different peptide substrates. As such, the generation of selective inhibitors may be feasible, though appropriate small molecule scaffolds must first be identified. Whether DPY19 inhibition, selective or otherwise, is of benefit in any therapeutic context remains an open-ended question. Developing the tools to probe these questions remains of great interest to our laboratory and the field more generally.

7.4 - Conclusions

The powerful combination of CRISPR-Cas9 technology and comparative cell-surface proteomics has provided the least biased insights yet into how the human *C*-mannosyltransferases DPY19L1 and DPY19L3 impact cell surface protein expression levels. The human monocytic cell line THP-1 was an ideal platform for these studies, as it is easily cultured and transfected, expresses a lot of putative *C*-mannosyltransferase substrate proteins, and can be easily differentiated into functionally distinct cell types. This data revealed that both DPY19L1 and DPY19L3 are important for the expression of a range of proteins, that they appear to be important to different subsets of proteins, and that more proteins require DPY19L1 in this cell lineage than DPY19L3. The data also alludes to the presence of C-mannosyltryptophan on a host of hitherto uninvestigated proteins, such as membrane transporters. Very large and significant decreases in the expression level of UNC5C and UNC5B mirror data presented in previous chapters for heterologous expression of UNC5. It is very likely that tryptophan mannosylation by DPY19L1 plays an important role in netrin signalling, which has important ramifications for development and a subset of Alzheimer's disease patients.

Chapter 8

Materials and Methods

8.1 - General procedures

8.1.1 - PCR

PCR reactions were carried out in 200 µl PCR tubes with a final reaction volume of 50 µl. Reactions contained 1 µl of template (≈1 ng/ml), 5 µl of each primer (5 µM), 14 µl dH_2O and 25 µl of Phusion polymer master mix (2x, ThermoFisher). Reactions were performed in a thermocycler (BioRad). then 10 µl DNA loading buffer (6x, NEB) was added to the reaction and the products purified on an agarose gel (0.5-1%) with SYBR safe DNA dye (Thermofisher) using electrophoresis (30 min, 90 V). DNA bands were visualised and excised using a blue light transilluminator. DNA was extracted from the agarose gel using the Wizard® SV Gel and PCR Clean-Up System (Promega).

8.1.2 - Cloning

Host plasmid vectors and dsDNA inserts were digested with the appropriate restriction enzymes (NEB) according to the manufacturer's protocols. The plasmid vector was also treated with recombinant shrimp alkaline phosphatase (rSAP, NEB) according to the manufacturer's protocols prior to purification by agarose gel electrophoresis (as described in 8.1.1.1). The digested dsDNA inserts were simply purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). Ligation reaction were performed using a Quick Ligation kit (NEB) according to the manufacturer's instructions. For transformation into bacterial cells, 2 µl of the ligation mixture was combined with 4 µl of KCM solution (5x) [201] and 14 µl of dH_2O. 20 µl of chemically competent TOP10 *E. coli* cells were added and the mixture was incubated on ice for 20 min followed by incubation at RT for 10 min. This 40 µl mixture was then plated onto LG agar plates with the appropriate antibiotic for the resistance marker used by the plasmid vector. Individual colonies were expanded, the plasmids isolated (Miniprep kit, Invitrogen) and Sanger sequencing (AGRF) performed to ensure that the plasmid possessed the desired DNA sequence.

8.1.3 - Integration of vectors into *Pichia pastoris*

pGAPZ and related plasmids (ThermoFisher) were integrated according to the manufacturer's protocols. Briefly, 10 µg of plasmid linearized with AvrII was transformed into 100 µl of *Pichia pastoris GS115* cells and plated onto YPDS (1% yeast extract, 2% peptone, 2% dextrose and 1 M sorbitol) plates supplemented with 100 µg/µl zeocin. The plates were incubated in the dark at 30 °C for 4 days until colonies could be observed. Individual clones were re-plated onto YPDS + 100 µg/µl zeocin with a scoring grid. Zeocin-resistant clones were picked into 1 ml YPD (1% yeast extract, 2% peptone, 2% dextrose) media and cultured at 30 °C overnight. Genomic DNA was isolated (see 8.1.4) and PCR was used to confirm integration of the plasmid into the relevant locus. Clones that were validated by PCR were then analysed by Western blot (see 8.1.7) with an appropriate antibody to confirm expression of the integrated gene.

pPIC9, pPIC9K and related plasmids (ThermoFisher) were integrated according to the manufacturer's protocols. Briefly, 10 µg of plasmid linearized with SalI or SacI was transformed into 100 µl of *Pichia pastoris* cells and plated onto histidine drop-out media plates. The plates were incubated at 30 °C for 4 days until colonies could be observed. Individual clones were picked into 1 ml YPD (1% yeast extract, 2% peptone, 2% dextrose) media and cultured at 30 °C overnight. Genomic DNA was isolated (see 8.1.4) and PCR was used to confirm integration of the plasmid into the relevant locus. Clones that were validated by PCR were cultured on small scale and protein expression induced with methanol (as described in the manufacturer's manual). Culture supernatants were evaluated by SDS-PAGE (see 8.1.6) and/or Western blot (see 8.1.7) with an appropriate antibody to confirm expression of the integrated gene.

8.1.4 - Isolation of *Pichia pastoris* genomic DNA

Yeast cell cultures were grown in YPD media (1 ml) at 30 °C and 220 rpm to an OD_{600} of 5.0. Cells were pelleted by centrifugation at 4 °C for 5 min at 1,500 × g. The cell pellet was washed twice with sterile H_2O (10 ml) by resuspension and pelleting by centrifugation. The pellet was resuspended in 200 µl of buffer containing 10 mM NaP_i @ pH 7.5, 1 M sorbitol, 10 mM EDTA, 10 mM DTT. The suspension was then treated with 3 units of lyticase for 60 min at 37 °C to digest the yeast cell wall. 200 µl of 1% SDS was added and the samples carefully inverted 6 times and incubated on ice for 5 min. 150 µl of 5 M potassium acetate @ pH 8.9 was added and the samples carefully inverted 6 times. The samples were then centrifuged at 10,000 × g

for 10 min at 4 °C. The supernatant was transferred to a fresh 1.5 ml microcentrifuge tube. 2 volumes of 100% ethanol were added to the supernatant and the samples incubated for 15 min at RT. The samples were centrifuged at 10,000 × g for 20 min at 4 °C. The supernatant was discarded and the pellet was resuspended in 200 µl of TE buffer containing 10 mM Tris @ pH 7.5, 1 mM EDTA. One volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the samples and the mixture vortexed vigorously before centrifugation at 16,000 × g for 5 min at RT. The upper aqueous phase was removed and placed in a fresh 1.5 ml microcentrifuge tube. 100 µl of 7.5 M NH_4OAc was added and the sample mixed. 750 µl of 100% ethanol was added and the sample mixed. The sample was incubated at -80 °C for at least 1 h. Centrifugation at 4 °C for 30 min at 16,000 × g, followed by careful removal of the supernatant provided the gDNA pellet. This was washed with 150 µl of 70% ethanol. The DNA pellet was air-dried in a fume hood for 1 h. The gDNA pellet was resuspended in sterile H_2O and stored at 4 °C.

8.1.5 – Fractionation of yeast sub-cellular compartments

Yeast cultures (100 ml) were grown in YPD to an OD_{600} of 2.0 then centrifuged at 1,500 × g for 5 min at 4 °C. The cell pellet was washed twice with filter-sterilised dH_2O then resuspended in 2 ml of buffer containing 30 mM Tris-HC1 pH 7.5, 50 mM KC1, 5 mM $MgCl_2$, 1 M sorbitol, 10 mM DTT and 20% v/v glycerol. Thirty units of lyticase was added and the cells incubated at 37 °C for 60 min. Protease inhibitor cocktail (Roche cOmplete) was added to the mixture and the cells were lysed using a French press, with the mixture being passed through the pressure cell 8 times. The lysate was centrifuged at 3,000 × g for 5 min at 4 °C. The supernatant was then centrifuged three times at 16,000 × g for 10 min at 4 °C g to remove mitochondria and cell nuclei. This supernatant was then centrifuged at 90,000 × g for 1 h at 4 °C. The pellet containing the yeast's microsomes was suspended in buffer containing 30 mM Tris-HC1 pH 7.5, 50 mM KC1, 5 mM $MgC1_2$, 1 M sorbitol, 10 mM DTT and 20% v/v glycerol and stored at 4 °C for no longer than 24 h.

8.1.6 - SDS-PAGE analysis

Protein samples containing SDS loading buffer (ThermoFisher) were heated at 95 °C for 2 min, cooled and loaded onto a NuPAGE™ Novex™ 4-12% Bis-Tris protein gel (ThermoFisher). Electrophoresis was performed in a gel tank using NuPAGE™ MES SDS Running Buffer (ThermoFisher) according to the manufacturer's instructions. Acrylamide gels were stained using a solution containing 0.2% Coomassie Brilliant Blue, 40% methanol, 10% acetic acid for

30 min at RT. De-staining was performed using a solution containing 10% methanol and 10% acetic acid for 4 h.

8.1.7 - Immunoblotting (Western blots)

Protein samples were subjected to SDS-PAGE as described in 8.1.6 but rather than being stained, proteins in the acrylamide gel were transferred to a nitrocellulose membrane using a mini blot module (ThermoFisher) according to the manufacturer's protocols (60 min and 10 V). The membrane was blocked with 4% skim milk in PBS for 60 min at RT. The membrane was washed with PBS + 0.1% tween 20, and probed with a primary antibody in PBS + 3% skim milk. The membrane was then washed with PBS + 0.1% tween 20 and probed with a secondary antibody conjugated to horse radish peroxidase (HRP) (ThermoFisher) for 60 min in PBS + 3% skim milk. The membrane was then washed with PBS + 0.1% tween 20, developed using the Immobilon Western Chemiluminescent HRP Substrate (Merck) and imaged using a ChemiDoc^{TM} Imaging System (BioRad).

8.1.8 - Expression of proteins in *Pichia pastoris* using methanol induction

A colony of the relevant protein-expressing yeast strain was used to inoculate 10 ml of YPD culture, which was grown for 24 h at 30 °C and 220 rpm. The 10-ml culture was used to inoculate 1 L of BMGY media (see Table 8.8) and this was cultured at 30 °C and 220 rpm until the OD_{600} reached 6.0 (≈24 h). The cells were gently pelleted by centrifugation at 1,500 × g for 20 min at 4 °C, then resuspended in BMMY media (see Table 8.8) and cultured at 30 °C and 220 rpm for 48 h, with an additional 0.25% v/v methanol being added after 24 h. The supernatant was harvested by centrifugation at 9,000 × g for 20 min at 4 °C. The supernatant was filtered (0.45 µm, Millipore), treated with a protease inhibitor cocktail (Roche cOmplete) and EDTA and NaN_3 added to final concentrations of 2 mM and 0.02%, respectively. The supernatant was concentrated to 100 ml using an Amicon stirred cell (Millipore) and membranes with an 8 kDa NMWL. This concentrate was dialysed against twice against 2 L of 50 mM Tris-HCl, 500 mM NaCl @ pH 7.5. After concentration and dialysis, this supernatant could be captured and purified by IMAC (see 8.1.9).

8.1.9 - Immobilized metal affinity chromatography (IMAC)

Immobilized metal affinity chromatography (IMAC) was performed using His GraviTrap TALON® columns (GE-Healthcare) according to the manufacturer's protocol. Briefly, the

Talon column was washed once with dH_2O then equilibrated into binding buffer (50 mM Tris-HCl, 500 mM NaCl @ pH 7.5). The concentrated and dialysed supernatant (see 8.1.8) was applied to the column under gravity. The column was washed with one column-volume of binding buffer. Protein was eluted from the column using elution buffer (50 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole @ pH 7.5). The purity of fractions from this column were assessed by SDS-PAGE (see 8.1.6). Protein fractions with desired protein were pooled and further purified by SEC (see 8.1.10) and/or IEX (see 8.1.11).

8.1.10 - Size exclusion chromatography (SEC)

Protein samples were purified by SEC using either Superdex 75 increase 10/300 GL or Superdex 200 increase 10/300 GL columns (GE-Healthcare) on an FPLC system. Unless otherwise stated, the buffer containing 50 mM Tris-HCl, 150 mM NaCl @ pH 7.5 was used for these purifications. Protein fractions were assessed for purity using SDS-PAGE (see 8.1.6).

8.1.11 - IEX

Protein samples in low-salt buffer were purified by IEX using either MonoQ 10/100 GL or MonoS 10/100 GL columns (GE-Healthcare) on an FPLC system. The buffers species, pH and salt gradients used varied from protein to protein and are detailed elsewhere. Protein fractions were assessed for purity using SDS-PAGE (see 8.1.6).

8.2 - Creating and studying *Pichia pastoris* strains harbouring DPY19

8.2.1 – Integrating CeDPY19 into *Pichia pastoris*

A synthetic gene encoding the *Caenorhabditis elegans* DPY19 (CeDPY19) protein with a C-terminal 3× Flag-tag was codon optimised for *Pichia pastoris* (IDT) and cloned into pGAPZ. This vector was integrated into *Pichia pastoris* using the general procedure outlined in 8.1.3. The sequence for the CeDPY19-$FLAG_3$ synthetic gene is as follows:

```
ATGGCTAAGAAGCCAAAGAACTCCCCAGAGAAGTCTAAGTACTCCTCCGATACTTCCTCCTCCTTGTACTCCCAA
ACTTGGTTGGCTTCCGTTGTTATCATCGGTTTGTTGGTTGGTTACATCAACTACCAGCACGTTTACACTTTGTTC
GAGAACGACAAGCACTTCTCCCACTTGGCTGACTTCGAAAGAGAGATGGCTTACAGAACTGAGATGGGTTTGTAC
TACTCCTACTACAAGACTATCATCAACGCTCCATCTTTCTTGGAGGGTGTTCAAGAGATCACTCACGACACTGTT
ACTGAACACGGTCACGAGATCAACACTTTGAACAGATTCAACTTGTACCCAGAGGTTATCTTGGCTTTCTTGTAC
AGACCATTCAGAGCTTTCGCTAAGTCCGCTAACTGGCAGATTGAGTTGTGTTGGCAGGTTAACAGAGGTGAGTTG
AGACCAGTTGAATCCTGTGAGGGTATTGGTAACCCACACTACTTCTACATCACTGGTGTTTTCATCGTTGCTGGT
```

ACTGTTGCTTCCTCCATCTTCTACTTGGGTGTTTTGGTTTCCGACTCCATTTTCGGTGGTTTCTTGTCCGTTTTG
TGTTTCGCTTTCAACCACGGTGAGGCTACTAGAGTTCAATGGACTCCACCATTGAGAGAGTCCTTCGCTTTCCCA
TTCATCATCGGTCACATTGCTATCTTGACTTTCGTTATCAAGTACAAAAAGTCCGGTCACTCCATGATCTTGTTG
TTGACTTCCATGGCTGTTCCAGCTTTGTTGTTCTGGCAGTTCACTCAGTTCGCTTTCTTCACTCAGATCTGTTCT
ATCTTTTTGGCTTTTTCCTTGGACTTGATCCCATTCTCCACTGCTAAGACTGTTATCCACTCCCACATCATCTCA
TTCTTGATCGGTTTTTTGTTGTTGTTCGGTAACGAGATGATGATCACTGCTCTTTACTTCCCTTCAATTTTAGCC
TTGGGAATGATCATTTATATTTCACCATTGCTTTCCAACCTGAAGTTCAGACCAGCTTACGTTTTGTTCTTGGCT
ATCATCTTCGCTTCCATCACTTTGGGTTTGAAGATCGGATTGTCCAAGGGTTTGGGTATTGAGGACGACGCTCAC
ATCTTCGACATCTTGAGATCCAAGTTCACTTCCTTCGCTAACTTCCACACTAGATTGTACACTTGTTCCGCTGAG
TTCGACTTCATCCAGTACTCCACTATCGAGAAGTTGTGTGGTACTTTGTTGATCCCTTTGGCTTTGATCTCCTTG
GTTACTTTCGTTTTTAACTTCGTTAAGAACACTAACTTGTTGTGGAGAAACTCCGAAGAGATCGGTGAGAACGGT
GAGATCTTGTACAATGTTGTTCAATTGTGTTGTTCCACTGTTATGGCTTTTTTGATCATGAGATTGAAGTTGTTC
ATGACTCCACACTTGTGTATCGTTGCTGCTTTGTTCGCTAACTCCAAGTTGTTGGGTGGTGACAGAATCTCCAAG
ACAATCAGAGTTTCCGCTTTGGTTGGTGTTATCGCTATCTTGTTCTACAGAGGTATCCCAAACATCAGACAGCAG
TTGAACGTTAAGGGTGAGTACTCCAACCCAGACCAAGAGATGTTGTTCGACTGGATTCAGCACAACACTAAGCAG
GACGCTGTTTTCGCTGGTACTATGCCAGTTATGGCTAACGTTAAGTTGACTACTTTGAGACCAATCGTTAACCAC
CCTCACTACGAACACGTTGGTATCAGAGAGAGAACTTTGAAGGTTTACTCCATGTTCTCCAAGAAGCCTATCGCT
GAGGTTCACAAGATTATGAAGGAAATGGGAGTTAACTACTTCGTTTTCCAGTTGATGAACTGTTCCAACGACGAG
AGAAGACCAGAGTGTGTTTACAGAGGAATGTGGGATGAAGAGGACCCTAAGAACTCTGGTAGAACAGCTTTGTGT
GACTTGTGGATCTTGGCTGCTAACTCTAAGGACAACTCCAGAATCGCTCCATTCAAGATCGTTTACAACGCTAAC
AGAAATTACATCGTTTTGAAGATTCTCGAGGACTACAAGGACCACGACGGTGATTACAAGGATCATGACATCGAC
TACAAAGATGACGACGACAAGTAA

8.2.2 - Generation of CeDPY19 mutants

Eight DPY19 mutants were made using the four primer PCR method. Primers for these reactions are provided in Table 8.1. PCR protocols are provided in 8.1.1. Mutant amplicons were cloned into pGAPZ using the EcoRI and NotI restriction sites. These vectors were integrated into *Pichia pastoris* using the general procedure outlined in 8.1.3.

Figure 8. 1: List of primers for making DPY19 mutants

Mutant	Reaction	Template	Forward primer	Reverse primer
E102A	PCR1	pGAPZ-DPY19	ACCTTTCCCAATTTTGGTTTCTCC	CAAAGTGTTGATCTCGTGACCGTGTGCAGTAACAGTGTCGTGAGTGATCTCTTGAAC
E102A	PCR2	pGAPZ-DPY19	GTTCAAGAGATCACTCACGACACTGTTACTGCACACGGTCACGAGATCAACACTTTG	GGCAAATGGCATTCTGACATCCTC
E102A	PCR3	PCR1 + PCR2	ACCTTTCCCAATTTTGGTTTCTCC	GGCAAATGGCATTCTGACATCCTC
E118A	PCR1	pGAPZ-DPY19	ACCTTTCCCAATTTTGGTTTCTCC	GTCTGTACAAGAAAGCCAAGATAACCGCTGGGTACAAGTTGAATCTGTTC
E118A	PCR2	pGAPZ-DPY19	GAACAGATTCAACTTGTACCCAGCGGTTATCTTGGCTTTCTTGTACAGAC	ACCTTTCCCAATTTTGGTTTCTCC

E118A	PCR3	PCR1 + PCR2	ACCTTTCCCAATTTTGG TTTCTCC	GGCAAATGGCATTCTG ACATCCTC
H375A	PCR1	pGAPZ-DPY19	ACCTTTCCCAATTTTGG TTTCTCC	GTGAACTTGGATCTCA AGATGTCGAAGATGGC AGCGTCGTCCTCAATA CCCAAACC
H375A	PCR2	pGAPZ-DPY19	GGTTTGGGTATTGAGGA CGACGCTGCCATCTTCG ACATCTTGAGATCCAAG TTCAC	ACCTTTCCCAATTTTG GTTTCTCC
H375A	PCR3	PCR1 + PCR2	ACCTTTCCCAATTTTGG TTTCTCC	GGCAAATGGCATTCTG ACATCCTC
E579A	PCR1	pGAPZ-DPY19	ACCTTTCCCAATTTTGG TTTCTCC	CAAAGTTCTCTCTCTG ATACCAACGTGTGCGT AGTGAGGGTGGTTAAC GATTGGTC
E579A	PCR2	pGAPZ-DPY19	GACCAATCGTTAACCAC CCTCACTACGCACACGT TGGTATCAGAGAGAGAA CTTTG	ACCTTTCCCAATTTTG GTTTCTCC
E579A	PCR3	PCR1 + PCR2	ACCTTTCCCAATTTTGG TTTCTCC	GGCAAATGGCATTCTG ACATCCTC
Y75F	PCR1	pGAPZ-DPY19	ACCTTTCCCAATTTTGG TTTCTCC	CGTTGATGATAGTCTT GTAGTAGGAGTAGAAC AAACCCATCTCAGTTC TGTAAGC
Y75F	PCR2	pGAPZ-DPY19	GCTTACAGAACTGAGAT GGGTTTGTTCTACTCCT ACTACAAGACTATCATC AACG	ACCTTTCCCAATTTTG GTTTCTCC
Y75F	PCR3	PCR1 + PCR2	ACCTTTCCCAATTTTGG TTTCTCC	GGCAAATGGCATTCTG ACATCCTC
R112A	PCR1	pGAPZ-DPY19	ACCTTTCCCAATTTTGG TTTCTCC	CCAAGATAACCTCTGG GTACAAGTTGAATGCG TTCAAAGTGTTGATCT CGTGACC
R112A	PCR2	pGAPZ-DPY19	GGTCACGAGATCAACAC TTTGAACGCATTCAACT TGTACCCAGAGGTTATC TTGG	ACCTTTCCCAATTTTG GTTTCTCC
R112A	PCR3	PCR1 + PCR2	ACCTTTCCCAATTTTGG TTTCTCC	GGCAAATGGCATTCTG ACATCCTC
R586A	PCR1	pGAPZ-DPY19	ACCTTTCCCAATTTTGG TTTCTCC	GGAGAACATGGAGTAA ACCTTCAAAGTTGCCT CTCTGATACCAACGTG TTCG
R586A	PCR2	pGAPZ-DPY19	CGAACACGTTGGTATCA GAGAGGCAACTTTGAAG GTTTACTCCATGTTCTC C	ACCTTTCCCAATTTTG GTTTCTCC
R586A	PCR3	PCR1 + PCR2	ACCTTTCCCAATTTTGG TTTCTCC	GGCAAATGGCATTCTGA CATCCTC
Y591F	PCR1	pGAPZ-DPY19	ACCTTTCCCAATTTTGG TTTCTCC	CGATAGGCTTCTTGGAG

				AACATGGAAAAGACCTTCAAAGTTCTCTCTCTGATACC
Y591F	PCR2	pGAPZ-DPY19	GGTATCAGAGAGAGAACTTTGAAGGTCTTTTCCATGTTCTCCAAGAAGCCTATCG	ACCTTTCCCAATTTTGGTTTCTCC
Y591F	PCR3	PCR1 + PCR2	ACCTTTCCCAATTTTGGTTTCTCC	GGCAAATGGCATTCTGACATCCTC

8.2.3 - Generation of IL12B mutants

Two IL12B mutants were made using the four primer PCR method. Primers for these reactions are provided in Table 8.2. PCR protocols are provided in 8.1.1. Mutant amplicons were cloned into pPIC9K using the EcoRI and NotI restriction sites. These vectors were integrated into *Pichia pastoris* using the general procedure outlined in 8.1.3.

Figure 8. 2: List of primers for making IL12B mutants

Mutant	Template	Forward primer	Reverse primer
W300Y	pGAPZ-DPY19	CTTTCATAATTGCGACTGGTTCC	AATTCGCGGCCGCTCAGGAACAGGGCACTGAGGCATATTCGGACCAGGAAGATGAATAG
W300F	pGAPZ-DPY19	CTTTCATAATTGCGACTGGTTCC	AATTCGCGGCCGCTCAGGAACAGGGCACTGAGGCAAATTCGGACCAGGAAGATGAATAG

8.2.4 - Yeast growth curves

A 5 ml YPD overnight culture (30 °C and 220 rpm) of *Pichia pastoris GS115* with and without CeDPY19 was used to inoculated 100 ml of YPD. These 100-ml cultures were grown at 30 °C and 220 rpm, with the OD_{600} recorded every 60 min for 15 h. The data was plotted using Prism 8 (GraphPad) and fitted to an exponential growth curve to obtain the growth exponent for each yeast strain.

8.2.5 - *In vitro* C-mannosylation assay

This assay was carried out in 1.5 ml microcentrifuge tubes with a final volume of 25 µl and contained 5 mM GDP-[^{3}H] Man (120 KBq), 5 mM GDP-Man, 100 mM MOPS (pH 7.5), 0.1% saponin, 2 mM $MnCl_2$, 2 mM $MgCl_2$, 2 mM ATP, 1 mM synthetic peptide and 2 µl of membrane vesicles (see 8.1.5). The mixture was incubated in a shaker at 400 rpm at 28 °C for 20 h. The reaction was stopped by adding 1 ml chloroform/methanol 3:2 (v/v) and 230 µl of

H_2O. The samples were vortexed briefly and centrifuged at 3000 × g for 5 min. 400 µl of the aqueous (upper) layer containing the peptide was diluted into 4 ml of dH_2O with 0.1% trifluoroacetic acid (TFA). This diluted aqueous solution was passed through a 200 mg C18 Solid Phase Extraction Cartridge (SepPak, Waters) using a disposable plastic syringe. The solid phase cartridge was washed two times with 10 ml dH_2O with 0.1% TFA. The labelled peptides were eluted with 4 ml of 100% methanol into a glass tube and the methanol evaporated. The residue was dissolved in 200 µl of methanol and 2 ml of scintillation solution (PerkinElmer) was added to the sample. This solution was transferred to counting tubes and beta-emission activity quantitated using a Microbeta scintillation counter (PerkinElmer).

8.2.6 - Immunoblotting for *Ce*DPY19

Clones that had positive integration of CeDPY19 into the *GS115* genome were grown in 1 ml of YPD media overnight at 30 °C at 220 rpm to an OD_{600} of 2.0. The culture was then centrifuged at 1500 × g for 10 min at 4 °C, cells were washed twice with sterile dH_2O and the pellet was resuspended in 200 µl of buffer containing 10 mM NaP_i @ pH 7.5, 1 M sorbitol, 10 mM EDTA, 10 mM DTT. Cell walls were digested using 5 units of lyticase for 60 min at 37 °C at 220 rpm to create spheroplasts. Protease inhibitor cocktail (Roche cOmplete) was added to the spheroplasts, which were then lysed using 200 µl of 10% SDS solution and 5 units of Benzonase® (Sigma) at 4 °C for 5 min with periodical inversion. 10 µl of 4x SDS loading buffer was added to 30 µl of this cell lysate and the mixture maintained at 37 °C for 10 min. It was crucial that these samples were not heated higher than this temperature, as this led to aggregation of the protein of interest. This sample was subjected to SDS-PAGE (see 8.1.6) and immunoblotting (see 8.1.7) using monoclonal ANTI-FLAG® M2 antibody (1/5000) (Sigma) as the primary antibody.

8.2.7 - Production of MBP-$FLAG_3$

The dsDNA sequence below was manufactured (IDT) and cloned (see 8.1.2) into the commercial pMAL-c5x (NEB) vector using the BamHI and XhoI restriction sites.

CCCTGGGATCCGACTACAAGGACCACGACGGTGATTACAAGGATCATGACATCGACTACAAAGATGACGACGACAAGTAAGCG
GCCGCCAGCTTGGGCCCGAACAAAAACTCATCTCAGAAGAGGATCTGAATTAAAAGCTTGACCATCACCACCATCACCATCTC
GAGTGAGT

This plasmid was transformed into chemically competent 'NEB Express' *E. coli* using the KCM method and plated onto LB-agar (100 µg/ml Amp) and incubated at 37 °C for 16 h. A

single colony was used to inoculate 10 ml of LB media containing 100 μg/ml Amp followed by incubation at 37 °C for 16 h. This culture was used to inoculate 1000 ml of "S-broth" (35 g tryptone, 20 g yeast extract, 5 g NaCl, pH 7.4) containing 100 μg/ml Amp, which was then incubated with shaking (250 rpm) at 37 °C until it reached an OD_{600} of 0.8. The culture was cooled to room temperature, IPTG added to a final concentration of 100 μM, and then incubated with shaking (200 rpm) at 18 °C for 19 h. The cells were harvested by centrifugation at 18,000 × g for 20 min at 4°C. The cell pellet was resuspended in 40 ml of binding buffer (50 mM Tris, 300 mM NaCl, 5 mM imidazole, pH 6.0) containing protease inhibitor cocktail (Roche cOmplete) and lysozyme (0.1 mg/ml) by nutating at 4°C for 30 min. Benzonase (1 μl) was added to the mixture then lysis was effected by sonication [10× (15 s on / 45 s off) at 45% amplitude]. The lysate was centrifuged at 18,000 × g for 20 min at 4°C and the supernatant collected. The supernatant was purified by IMAC (see 8.1.9) and SEC on a Sephacryl 200 column (see 8.1.10).

8.3 - Producing and characterising *C*-mannosylated proteins in yeast.

8.3.1 - Reporter protein constructs

DNA sequences encoding human IL21R (20-232 aa), EPOR (25-250 aa), PLR (25-234 aa) and IL12B with a hexahistidine tag, as well as sequences encoding human RNaseII, mindin (Spondin 2) TSR domain (277-331 aa) and *C. elegans* UNC5 TSR2 domain (274-326 aa) with a SUMO fusion partner and hexahistidine tag, were codon harmonised for *Pichia pastoris* (Table 8.3), synthesised (IDT), and cloned into pPIC9K using the NotI and XhoI restriction sites. Sanger sequencing-validated plasmids were integrated into *P. pastoris* strains (see 8.1.3). Proteins were expressed and purified described above (see 8.1.8-10).

Figure 8. 3: DNA sequence of reporter proteins

Protein	Synthetic dsDNA cloned into pPIC9K @ EcoRI/NotI
Human IL21R ectodomain	ACGTAGAATTCTGTCCTGACCTGGTTTGCTATACTGATTACTTGCAAACAGTCATTTGCATTC TTGAAATGTGGAATCTTCACCCTTCCACACTTACACTTACGTGGCAAGACCAGTATGAAGAGT TGAAGGACGAAGCAACAAGTTGTTCTTTGCACAGATCCGCCCATAATGCTACTCACGCCACTT ATACTTGTCATATGGATGTTTTCCATTTCATGGCCGACGACATTTTCTCCGTGCAAATCACAG ACCAGTCTGGGCAGTATAGTCAGGAATGCGGTTCCTTTTTGTTGGCAGAATCTATAAAACCAG CTCCACCATTCCAGGTGACAGTTACCTTCTCAGGACAATACCAAATCTCCTGGCGTTCAGATT ATGAAGACCCAGCCTTCTACATGCTAAAAGGTAAGTTACAATACGAGCTGCAATATCGTAATC GAGGAGACCCCTGGGCTGTCTCACCAAGAAGGAAGCTGATCAGTGTCGATTCTAGATCCGTGT CTTTGTTACCTTTAGAGTTTAGAAAGGATTCCTCCTATGAGTTACAAGTTAGAGCTGGTCCAA

	TGCCAGGTAGTTCTTACCAAGGTACCTGGAGTGAATGGAGTGATCCTGTTATTTTCCAAACTC AATCCGAGGAGTTGAAAGAACATCATCACCATCACCATTAAGCGGCCGCGAATT
Human EPOR ectodomain	GAGACGAATTCCCTCCTCCTAATTTACCAGACCCAAAGTTTGAATCAAAAGCTGCTTTATTAG CTGCTAGAGGACCTGAGGAGCTGCTATGTTTTACCGAGAGATTAGAGGACCTAGTATGCTTCT GGGAAGAAGCTGCATCCGCTGGTGTAGGCCCCGGTAATTACAGTTTCTCCTACCAATTGGAAG ATGAGCCTTGGAAATTGTGTAGATTGCATCAAGCACCAACAGCTAGAGGTGCTGTTAGATTCT GGTGTTCTCTGCCAACAGCTGACACATCTTCATTCGTACCTTTGGAATTACGTGTCACAGCAG CATCTGGTGCTCCAAGATATCATCGAGTTATTCATATCAATGAGGTTGTGCTTTTGGATGCTC CCGTTGGACTAGTGGCAAGATTGGCAGATGAATCTGGTCACGTTGTGCTAAGATGGCTACCTC CACCAGAAACTCCTATGACTTCACATATCAGGTATGAAGTAGACGTTTCAGCCGGTAATGGTG CAGGATCTGTTCAGAGAGTGGAAATTTTAGAGGGACGTACTGAGTGCGTATTGTCTAATTTGA GGGGAAGGACCAGATACACCTTCGCCGTTCGTGCTCGAATGGCTGAACCATCCTTTGGCGGTT TCTGGTCTGCATGGTCTGAGCCCGTGTCCCTTTTGACTCCTTCTGACCTAGATCATCATCACC ATCACCATTAAGCGGCCGCGAATT
Human PLR ectodomain	ACGTAGAATTCCAACTGCCACCCGGAAAGCCCGAAATTTTCAAGTGCAGGAGTCCCAATAAGG AAACGTTCACCTGCTGGTGGAGGCCCGGCACAGATGGTGGTTTGCCCACAAACTACTCTCTGA CTTATCACAGGGAAGGCGAGACATTAATGCATGAATGCCCAGACTACATTACAGGTGGACCTA ATTCTTGCCATTTTGGCAAGCAGTATACCTCAATGTGGAGGACATACATTATGATGGTGAATG CTACTAACCAAATGGGATCTTCTTTCTCAGATGAACTTTACGTCGATGTCACCTACATTGTTC AACCAGATCCCCCTCTTGAACTAGCAGTGGAGGTTAAACAACCAGAGGACCGTAAACCTTATC TGTGGATCAAGTGGTCCCCACCAACTCTGATAGACTTAAAAACGGGGTGGTTCACCCTTCTGT ACGAAATAAGATTGAAGCCTGAGAAAGCAGCAGAATGGGAAATTCACTTCGCAGGACAACAAA CCGAGTTCAAGATCCTGAGTTTGCATCCAGGCCAGAAGTACTTGGTACAAGTTAGATGTAAGC CAGATCACGGATATTGGTCCGCCTGGTCACCAGCTACCTTTATTCAGATTCCAAGTGACTTCA CGATGAACGACCACCACCATCACCATCATTAGGCGGCCGCGAATT
Human IL12B	GAGACGAATTCCATCACCATCATCATCACATCGAAGGTAGAATCTGGGAGCTGAAGAAAGACG TTTACGTTGTCGAACTAGATTGGTATCCTGATGCACCTGGTGAAATGGTCGTTCTAACTTGTG ATACACCTGAAGAGGACGGAATAACTTGGACTCTGGATCAGTCTTCCGAAGTTCTAGGGTCTG GTAAAACTCTGACTATCCAAGTCAAAGAGTTTGGAGACGCTGGCCAATATACTTGTCATAAGG GTGGGGAGGTGTTATCCCACTCTTTGCTATTGTTGCACAAAAAAGAAGATGGAATTTGGTCTA CTGATATTTTGAAGGATCAAAAAGAGCCAAAGAACAAGACCTTTTTGAGATGCGAAGCTAAAA ACTATTCCGGTCGTTTCACCTGTTGGTGGTTGACGACCATCTCCACAGACTTGACTTTCTCTG TTAAGTCATCTCGTGGTAGTTCTGATCCACAAGGAGTCACTTGCGGTGCCGCTACCTTGTCAG CCGAACGTGTTAGAGGTGATAACAAGGAATATGAGTATTCTGTTGAATGTCAAGAAGATTCAG CATGTCCTGCAGCTGAAGAGTCTCTTCCAATCGAGGTCATGGTTGATGCAGTACACAAGTTGA AGTACGAAAACTACACTTCTTCTTTTTTCATTCGTGACATTATCAAACCCGATCCCCCTAAGA ATTTGCAGTTGAAGCCTTTAAAAAATTCTAGACAAGTGGAGGTTTCTTGGGAATACCCTGACA CTTGGTCTACTCCCCATTCTTATTTTTCTCTAACTTTTTGTGTACAAGTTCAAGGCAAGTCTA AGCGAGAGAAGAAGGATCGAGTCTTCACTGACAAGACTAGCGCAACAGTCATTTGTAGAAAGA ACGCATCAATTTCTGTTAGGGCTCAGGACCGTTACTATTCATCTTCCTGGTCCGAATGGGCCT CTGTGCCCTGTTCCTGAGCGGCCGCGAATT
SUMO-Human mindin TSR fusion	ACGTAGAATTCGACTACAAAGACGATGACGACAAGGGACATCACCATCATCATCACGGAGGTT CGGACTCAGAAGTCAATCAAGAGGCTAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTC ACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTC CTTTAAGAAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAACGT TCTTGTACGACGGTATTGAAATTCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATA ACGATATTATTGAGGCTCACCGCGAACAGATTGGAGGTACTAGTGATTGCGAGGTTTCCTTGT GGTCTTCATGGGGTTTGTGCGGTGGACACTGCGGCAGATTGGGCACAAAGTCCAGGACTCGTT ATGTTAGAGTGCAACCAGCAAACAACGGAAGTCCCTGCCCTGAGTTGGAAGAGGAGGCAGAAT GTGTTCCTGACAACTGTGTTGGATCTTGGTCACACCCACAGTTTGAGAAATAAGCGGCCGCGA ATT
SUMO-*C. elegans* UNC5 TSR fusion	ACGTAGAATTCGACTACAAAGACGATGACGACAAGGGACATCACCATCATCATCACGGAGGTT CGGACTCAGAAGTCAATCAAGAGGCTAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTC ACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTC CTTTAAGAAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAACGT TCTTGTACGACGGTATTGAAATTCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATA ACGATATTATTGAGGCTCACCGCGAACAGATTGGAGGTACTAGTCCCTGCAAGCTGGATGGTG GTTGGTCTTCTTGGTCTGACTGGTCAGCATGTTCCTCTTCTTGTCACCGATACAGGACCAGAG CATGCACTGTCCCCCCTCCAATGAACGGAGGTCAACCATGTTTTGGGGATGATTTGATGACTC AGGAATGTCCAGCTCAGTTGTGGTCTCACCCACAATTCGAAAAGTAAGCGGCCGCGAATT
SUMO-human RNaseII fusion	ACGTAGAATTCGACTACAAAGACGATGACGACAAGGGACATCACCATCATCATCACGGAGGTT CGGACTCAGAAGTCAATCAAGAGGCTAAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTC ACATCAATTTAAAGGTGTCCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTC CTTTAAGAAGGCTGATGGAAGCGTTCGCTAAAAGACAGGGTAAGGAAATGGACTCCTTAACGT TCTTGTACGACGGTATTGAAATTCAAGCTGATCAGACCCCTGAAGATTTGGACATGGAGGATA ACGATATTATTGAGGCTCACCGCGAACAGATTGGAGGTACTAGTGAAAACTTGTACTTCCAAG GTAAGCCACCACAATTCACTTGGGCTCAATGGTTCGAAACTCAACACATCAACATGACTTCTC AACAATGTACTAACGCTATGCAAGTTATCAACAACTACCAAAGAAGATGTAAGAACCAAAACA CTTTCTTGTTGACTACTTTCGCTAACGTTGTTAACGTTTGTGGTAACCCAAACATGACTTGTC

	CATCTAACAAGACTAGAAAGAACTGTCACCACTCTGGTTCTCAAGTTCCATTGATCCACTGTA ACTTGACTACTCCATCTCCACAAAACATCTCTAACTGTAGATACGCTCAAACTCCAGCTAACA TGTTCTACATCGTTGCTTGTGACAACAGAGACCAAAGAAGAGACCCACCACAATACCCAGTTG TTCCAGTTCACTTGGACAGAATCATCTGAGCGGCCGCGAATT

8.3.2 - Reduction and alkylation of protein samples for proteolysis

Recombinant protein samples were mixed in buffer containing 2 M GnHCl, 5 mM DTT at 55 °C for 30 min. Samples were cooled to RT and treated with 5 mM iodoacetamide for 30 min with the exclusion of light. The reaction was quenched by the addition of 10 mM DTT and incubated at RT for 5 min. Nine volumes of ice-cold ethanol was added to the samples, which were mixed well and incubated at -20 °C overnight. Samples were centrifuged for 15 min at 13,000 × g at 4 ^{0}C. the supernatant was discarded and 1 ml of ice-cold ethanol was added to the pellet and incubated at -20 ^{0}C for 60 min. The samples were then centrifuged at 13,000 × g for 15 min at 4 °C. The supernatant was discarded and the protein pellet air-dried in a fume-hood.

8.3.3 - Proteolysis of proteins for mass spectrometry

Reduced and alkylated samples (see 8.3.2) were resuspended in 40 mM NH_4HCO_3 buffer and treated with GluC (Promega) or trypsin (Promega) for 24 h (GluC) or 16 h (trypsin) at 37 °C. Peptides were captured using C_{18} stage tips [202, 203], eluted with buffer containing 0.1% formic acid and 80% MeCN, evaporated and stored at -20°C prior to analysis by LC-MS.

8.3.4 - Identification of glycopeptides using reversed phase LC-MS, CID MS-MS and HCD MS-MS

Purified peptides were re-suspended in Buffer A* (0.1% TFA, 2% MeCN) and subjected to analysis on either a Q-Exactive or Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (ThermoScientific). Q-Exactive analysed samples were separated using an in-house packaged 25 cm, 75 µm inner diameter, 360 µm outer diameter, 1.7uµm 130 Å CSH C_{18} (Waters) reverse phase analytical column with an integrated HF etched nESI tip. Samples were loaded directly onto the column using an ACQUITY UPLC M-Class System (Waters) at 600 nl /min for 20 min with Buffer A (0.1% FA, 2% MeCN) and eluted at 300 nl/min using a gradient altering the concentration of Buffer B (0.1% FA, 80% MeCN) from 2% to 32% B over 60 mins, then from 32% to 40% B in the next 10 min, then increased to 80% B over 8 min period, held at 100% B for 2 min, and then dropped to 2% B for another 10 min. RP separated peptides were

infused into the Q-Exactive and data acquired using data dependent acquisition with one full precursor scan (resolution 70,000; 440-2000 m/z, AGC target of 1×10^6) followed by 10 data-dependent HCD MS-MS events (resolution 35k AGC target of 2×10^5 with a maximum injection time of 150 ms, NCE stepping set to 20; 26; 32) with 60 s dynamic exclusion enabled. Fusion™ analysed samples were separated using a two-column chromatography set up comprising a PepMap100 C18 20 mm x 75 µm trap and a PepMap C18 500 mm x 75 µm analytical column (ThermoScientific). Samples were concentrated onto the trap column at 5 µl/min for 5 min and infused into an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer at 300 nl/min via the analytical column using a Dionex Ultimate 3000 UPLC (ThermoScientific). 95 min gradients were run altering the buffer composition from 1% buffer B to 28% B over 60 min, then from 28% B to 40% B over 10 mins, then from 40% B to 100% B over 2 min, the composition was held at 100% B for 3 min, and then dropped to 3% B over 5 mins and held at 3% B for another 15 min. The Lumos™ Mass Spectrometer was operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (120,000 resolution) every 3 s and Orbitrap EThcD, Orbitrap HCD and ion trap CID scans for each selected precursor (maximum fill time 100 ms, AGC $1*10^5$ and with a resolution of 30,000 for Orbitrap MS-MS scans).

8.3.5 - Mass spectrometry data analysis

Identification of proteins and *C*-glycosylated peptides was accomplished using MaxQuant (v1.5.3.1) [204]. Searches were performed against the predicted amino acid sequence of each protein with carbamidomethylation of cysteine set as a fixed modification. Searches were performed with trypsin cleavage or GluC specificity based on the digestion type allowing 2 miss cleavage events and the variable modifications of oxidation of methionine, Hexose (W) ($C_6H_{10}O_5$ addition to tryptophan) and acetylation of protein N-termini. The precursor mass tolerance was set to 20 parts-per-million (ppm) for the first search and 10 ppm for main search, with a maximum false discovery rate (FDR) of 1.0% set for protein and peptide identifications. The resulting outputs were processed within the Perseus (v1.4.0.6) [205] analysis environment to remove reverse matches and common proteins contaminates prior to further analysis. Glycopeptides identified were manually assessed according to the guidelines of Chen *et al* [206], annotated manually with the aid of Protein Prospector tool MS-Product (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct) and are provided within the supplementary figures. Occupation analysis was undertaken using extracted ion

chromatograms of identified peptide species. Extracted ion chromatograms of the monoisotopic peak, ±10ppm, of identified glycopeptides and unmodified peptides species were generated within Xcalibur (version 2.2, Thermo Scientific). Peaks were processed with a 15-point Gaussian smooth and the area under the curve calculated. The resulting areas were used to compare peptide species with occupation rates provided as a percentage of the total ion current of compared species.

8.3.6 - Differential scanning fluorimetry

Differential scanning fluorimetry was performed in 0.1 ml strip tubes with a final reaction volume of 20 µl. Briefly, 1-3 µg of protein was mixed with buffer containing 50 mM Tris-HCl, 150 mM NaCl @ pH 7.5 along with 1 µl of SYPRO™ Orange Protein Gel Stain (1:100 in DMSO). The tubes were then placed in a Rotor-Gene Q and samples were heated from 25 °C to 80 °C in 1 min intervals. The raw data was then processed by truncating the data set to centre on the sigmoidal melting curve. This data was normalised and a Boltzmann Sigmoidal curve was fitted to provide a calculated T_m value. All experiments were conducted in triplicate.

8.3.7 - RNaseII activity assay

The RNase activity assay was adopted and modified from Tripathy et al [157] for RNaseII. Briefly, the reaction was conducted in 1 ml disposable cuvettes with reaction volumes of 1000 µl in buffer containing 50 mM Tris, 150 mM NaCl at pH 7.5 at 37 °C. RNA concentrations of 60, 50, 40, 30, 20, 16, 8, and 4 µg/ml were incubated with 4 µM ethidium bromide for 30 min at RT. RNaseII-GS115 and RNaseII-DPY19 at 0.004 mg/ml was added to each sample and the decay of florescence emission at 600 nm was continuously monitored for 10 min at 30 °C. The sample was excited at 510 nm. The data from 0-60 s fit well to a linear regression model, with the slope providing the initial velocity of the enzyme under that condition. These initial rates, were plotted with respect to substrate concentration and fit with a Michaelis-Menten model using Prism 8 (GraphPad) to provide the kinetic parameters for the different glycoforms of RNaseII.

8.3.8 - ELISA assay for LPS binding by mindin

The ELISA assay for LPS binding by mindin was adapted and modified from He *et al.* [81]. Briefly, 100 µl of 10 µg/ml of casein, mindin-GS115 and mindin-DPY19 diluted in buffer containing 100 mM sodium bicarbonate at pH 9.6 were coated on Nunc-Immuno MicroWell

96-well plates (ThermoFisher) overnight at 4 °C. Wells were washed four times with PBS + 0.2% tween-20, blocked with 100 µL PBS + 3% BSA at RT for 1 h, then washed four times with PBS + 0.2% tween-20. 100 µl of LPS was incubated in wells at concentrations of 300 µg/ml, 100 µg/ml, 10 µg/ml, 1 µg/ml, 100 ng/ml and 0 ng/ml at RT for 1 h. These wells were then washed four times with PBS + 0.2% tween-20 and probed using 20 µL of 1 µg/ml (in PBS+ 3%BSA) of anti-LPS antibody (Biomatik) at RT for 2 h. Four washes with PBS + 0.2% tween-20 came next, followed by incubation with 20 µL of 1 µg/ml (in PBS+ 3%BSA) goat anti-mouse IgG-HRP conjugate (ThermoFisher) at RT for 1 h. The wells were then washed four times with PBS + 0.2% tween-20, then 100 µl of 1 mg/ml OPD (o-Phenylenediamine dihydrochloride) (ThermoFisher) developing reagent was incubated in wells for 20 min. The reaction was stopped with sulfuric acid and absorbance was measured at 490 nm. Absorbance data was plotted using Prism 8 (GraphPad).

8.4 – Mammalian cell culture, cell line generation and cell surface proteomics.

8.4.1 - THP-1 cell culture

THP-1 cells were maintained in Roswell Park Memorial Institute (RPMI 1640) medium supplemented with 10% fetal calf serum and 100 µg/ml penicillin- streptomycin at 37 °C with 5% CO_2. Cells were sub-cultured in fresh media regularly to maintain cell viability and numbers.

8.4.2 - HEK293T cell culture

HEK293T cells were maintained in DMEM media supplemented with 10% fetal calf serum and 100 µg/ml penicillin- streptomycin until 90% confluency and split 1:5 to 1:10 at 37 °C with 5% CO_2 to maintain cell viability and numbers.

8.4.3 - Lentivirus production/transfection for CRISPR-Cas9 experiments

The day before transfection, 3 x 10^6 HEK293T cells were seeded into a 10-cm dish in DMEM media supplemented with 10% fetal calf serum and 100 µg/ml penicillin- streptomycin and incubated at 37 °C with 5% CO_2 overnight. On the day of transfection, the transfection mix was prepared by adding 4 µg of CRISPR guide RNA vector, 2 µg PMDL vector, 1 µg of pRSV-Rev vector, 1 µg VSVG vector, 24 µl Fugene 6 and 300 µl optimum media and incubated for

20 min at RT. The media for the cells was replaced with fresh media and transfection mix was added to the cells dropwise and they were incubated at 37 °C overnight. The next day, the media was removed from the cells and 6 ml of fresh media was added to the cells and incubated overnight at 32 °C. Target cells for transfection with lentivirus were seeded at 100,000 cells per well in a 6-well plate. The next day, the lentivirus was harvested and filtered using a 0.45 µM filter and 6 µl of polybrene was added to the virus. Spin transfection was initiated by adding 3 ml of lentivirus and cells were centrifuged at 3000 × g for 60 min at 32 °C. the cells were then incubated at 37 °C for 4 h and the media was then replaced and then incubated at 37 °C overnight. After 24 h, the media for the infected cells was changed and cells were split in transferred into a T25 flask. The cells were allowed to rest for 48-72 hours and viability was assessed. Cas9 was then induced using doxycycline (50 mg/ml; diluted to 1/10,000) for 4 days. After doxycycline treatment, cells were allowed to rest for 4 days with two media changes. Cells were then sorted using FACS gated for mCherry (Cas9) and GFP/CFP (guide RNA) into 96-well plates in 100 µl of fresh media. After 10 days, the cells were transferred into 24-well plates with 300 µl of media. After the cells were confluent, 16 clones were selected for genomic DNA extraction (see 8.4.5), PCR amplification and sequencing (see 8.4.6).

8.4.5 - Isolation of genomic DNA from THP-1 cells

Five million cells were centrifuged at 250 × g for 5 min. Genomic DNA was isolated from the cell pellet using the PureLink® Genomic DNA Mini Kit (ThermoFisher) according to the manufacturer's instructions. gDNA was stored at 4 °C.

8.4.6 - PCR amplification and sequencing of CRISPR'd genomic DNA

PCR reactions were carried out in 200 µl PCR tubes with a final reaction volume of 50 µl containing 1 µl of genomic DNA, 5 µl of each primer (5 µM stock, Table 7.4), 14 µl dH_2O and 25 µl of GoTAQ (Promega). Reactions were performed in a thermocycler (BioRad), 10 µl of 6x DNA loading dye was added, and the sample subjected to electrophoresis (30 min at 90 V) on a 1% agarose gel containing SYBR safe DNA dye. Bands corresponding to the PCR product were excised and the DNA extracted from the agarose gel using the Wizard® SV Gel and PCR Clean-Up System (Promega). These amplicons were subjected to Sanger sequencing (AGRF) for each of the *dpy19l1* and *dpy19l3* knockout and double-knockout clones.

Figure 8. 4: Sequencing primers for CRISPR of *dpy19l1* and *dpy19l3*

Gene	Product size (bp)	Forward primer	Reverse primer
dpy1911	560	GTTCATGGTGAGAGAGAG	GGAGCATCTACTCTTTCTAC
dpy1913	645	AGTAAGGCTAACCTGGGCTTG AAC	GCAGAAATTGTGTGTGGGTTTGC

8.4.7 - Sample preparation for cell surface proteomics

Cells were harvested at 250 × g for 5 min at 4 °C and washed twice with ice-cold PBS (pH 7.4). The sialic acid residues of cell surface proteins were oxidised and biotinylated with 1 mM sodium meta-periodate (ThermoFisher), 200 μM amino-oxy-biotin (Biotium) and 10 mM aniline (Sigma) in PBS at pH 6.7 for 1 h at 4 °C, with cells at a concentration of $2x10^7$ cells/ml. The oxidation and labelling reaction was quenched by the addition of 1 mM glycerol. Cells were washed three times with PBS at pH 7.4, then harvested by centrifugation at 400 × g for 5 min at 4 °C. The cell pellet was resuspended in lysis buffer containing 1% Triton X-100, 150mM NaCl, protease inhibitor cocktail (Roche, cOmplete), 10 mM Tris-HCl, pH 7.6 and lysed for 60 min at 4 °C. Nuclei were removed by centrifugation at 2800 × g for 10 min at 4 °C, then 16,000 × g for 10 min at 4 °C. The lysate was incubated with 50 μl of high-affinity streptavidin-agarose for 1 h 4 °C. The agarose beads were washed twice with lysis buffer, followed by washing three times with PBS + 0.5% (w/v) SDS. 100 mM DTT in 200 μl PBS + 0.5% SDS was added to the beads and incubated for 20 min at RT. The beads were then washed with UC buffer (6 M urea, 100 mM Tris-HCl, pH 8.5) 10 times. 50 mM iodoacetamide in 100 μl of UC buffer was added to the beads and they were incubated for 20 min at RT in the dark. The beads were then washed 10 times with UC buffer, followed by washing with PBS four times and dH_2O three times. Proteins were digested on-bead with 2 μg Trypsin Gold in 100μl 50 mM NH_4HCO_3 for 16 h. The samples were centrifugation at 1000 × g for 1 min and the supernatant containing the peptides of interest collected for mass spectrometry analysis (performed by Dr. Jarod Sandow at WEHI).

8.4.8 - PMA differentiation

1.5×10^6 cells were seeded into 100 mm dish containing 5 ml of RPMI 1640 media supplemented with 10% fetal calf serum and 100 μg/ml penicillin-streptomycin. PMA was added to a final concentration of 10 ng/ml and the cells were incubated at 37 °C with 5% CO_2 for 48 h. These cells were then washed once with PBS and 5 ml of the same media was added

to the cells, which were incubated at 37 °C with 5% CO_2 for an additional 24 h. These adherent cells were then washed three times with PBS, trypsinised to detach cells, and then re-seeded into a fresh 100 mm dish containing 5 ml of the same media. Once these cells had reached satisfactory confluence they were harvested using EDTA-mediated detachment and prepared for cell surface proteomics (see 8.4.8).

8.5 – Antibody generation and characterisation.

8.5.1 - Proteolysis and purification of IL21R

Before the proteolysis of IL21R, clostripain required activation for 1 h at RT. The clostripain activation reaction included 100 µg of clostripain in 10 mM Tris-HCl buffer with 2.5 mM DTT and 1 mM $CaCl_2$. The proteolysis reaction included 5 mg of IL21R, 10 mM Tris-HCl buffer with 2.5 mM DTT, 1 mM $CaCl_2$ and the addition of the entire clostripain activation reaction. After 2 h of digestion at 37 °C, protease inhibitor cocktail (Roche cOmplete) was added to the reaction and the sample was stored on ice before IMAC purification. The *C*-mannosylated peptide of interest was then purified with IMAC using a 1 ml HisTrap HP column (GE-Healthcare). The HisTrap HP column was equilibrated with binding buffer (50 mM Tris-HCl, 500 mM NaCl @ pH 7.5) and the proteolysis reaction applied to the column. The column was then washed with binding buffer for a total of 10 column volumes then eluted using elution buffer (50 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole @ pH 7.5). Protein fractions were assessed via SDS-PAGE for the peptide of interest. The fraction containing the peptide of interest was further purified by SEC using a Superdex 75 increase 10/300 GL column (GE-Healthcare) equilibrated with 50 mM Tris-HCl, 150 mM NaCl @ pH 7.5. The eluted protein fractions were assessed by SDS-PAGE and the fractions containing the peptide of interested were pooled and purified further using high performance liquid chromatography (HPLC). HPLC was performed on a C8 semi-prep column equilibrated with buffer A (95% H_2O, 5% MeCN and 0.1% TFA), the sample loaded and then eluted with a gradient from 5-100% of buffer B (100% MeCN and 0.1% TFA) across 12 column volumes (30 mL). Fractions were analysed using LC-MS to identify the peptide and assess its purity. Fractions containing pure *C*-mannosylated peptide were pooled and lyophilised, then stored at -20 °C.

8.5.2 - Conjugation of peptide to maleimide activated KLH and BSA

The purified *C*-mannosylated peptide was dissolved in buffer (50 mM NaPi, 150 mM NaCl, 5 mM EDTA, at pH 8) and Traut's reagent added (final concentration 700 μM) followed by incubation for 1 h at 22 °C. This thiolated peptide was rapidly purified using a PD MiniTrap G-10 desalting column (GE-Healthcare) into conjugation buffer (100 mM NaPi, 150 mM NaCl, 5 mM EDTA, pH 7.2). For conjugation to KLH, 600 μg of maleimide-activated KLH was combined with 400 μg of the desalted thiolated peptide in conjugation buffer and the sample incubated at 22 °C for 2 h. Reduced glutathione was added (final concentration 150 μM) and incubated for 20 min at 22 °C to quench the reaction. The sample was then desalted into DMPBS using a Zeba spin column (ThermoFisher). For conjugation to BSA, 400 μg maleimide-activated BSA was added to 300 μg of the desalted thiolated peptide in conjugation buffer and the sample incubated at 22 °C for 2 h. Cysteine was added (final concentration 200 μM) and incubated for 20 min at 22 °C to quench the reaction. The sample was desalted into storage buffer (80 mM NaPi, 900 mM NaCl, 100 mM sorbitol) using a Zeba spin column (ThermoFisher).

8.5.3 - Antigen immunisation schedule for mice

At day 0, four mice were immunised with 45 μg of peptide-KLH conjugate with Freund's complete adjuvant. At day 28, the mice were boosted with 45 μg of peptide-KLH conjugate with Freund's incomplete adjuvant. This was repeated at day 56. At day 68, the mice were bled and serum antibody titres against peptide-BSA conjugate assessed using ELISA and western blot. The mice were boosted with 15 μg of peptide-KLH conjugate for a pre-fusion boost at day 84. At day 88 the spleens were harvested for hybridoma generation, which was performed by the WEHI monoclonal antibody facility using established protocols [172, 207, 208]

8.5.4 - SPR

Surface plasmon resonance experiments were performed on a Biacore 4000 using a CM5 sensor chip (GE-Healthcare). Peptide was immobilised onto the CMC surface using EDC coupling. Antibodies were flown over the surface at the concentrations described in Chapter 5. The sensorgrams were fit using the bivalent binding model.

8.5.5 - Papain digestion of IgG to generate F_{ab} domains

F_{ab} fragments were generated through papain digestion of the 5G12 monoclonal antibody. Papain (Sigma) was prepared at 4 mg/mL in 100 mM potassium phosphate buffer, pH 6.1 and was activated by the addition of 2.5 mM EDTA, and 5 mM DTT. The activated papain was diluted to 0.05 mg/mL with 100 mM KP_i buffer, pH 6.1, to ablate the effect of the reducing agent on the monoclonal antibody. Monoclonal antibody was prepared at 2 mg/mL in 100 mM potassium phosphate buffer, pH 6.1, and the enzymatic digestion was initiated by mixing equal parts monoclonal antibody with activated papain. The enzymatic digestion was carried out for 60 min at 37 °C. The reaction was halted by the addition of $1/10^{th}$ the volume of 3 M Tris-HCl, pH 8.5, and freshly prepared iodoacetamide to 30 mM. To remove Fc domains and undigested monoclonal antibody, the digest was passed through a column of protein G immobilised on agarose (ThermoFisher). The flow-through and wash fractions were collected and analysed by SDS-PAGE. The protein G flow-through contained near-pure F_{ab}. This was further purified by SEC (Superdex 200 HiLoad 16/600, GE Healthcare – see 8.1.10) pre-equilibrated in 25 mM Tris-HCl, 150 mM NaCl, pH 7.4. The F_{ab} eluted as a single peak and its purity was assessed by SDS-PAGE before pooling relevant fractions. A 10kDa NMWL centricon filter (Millipore) was used to concentrate the purified F_{ab} to 15 mg/mL as determined by A_{280}.

8.5.6 - Crystallization of the F_{ab}–peptide complex and structure determination

The F_{ab}–peptide crystal was grown over 42 days in sitting drops at 20 °C by mixing 150 nL well solution containing 10% (v/v) 2-propanol, 20% (w/v) polyethylene glycol 4000, 0.1 M sodium HEPES, pH 7.5, with 150 nL F_{ab}–peptide solution at a 1:1.2 molar ratio at 10 mg/mL. The crystal was flash frozen in mother liquor using liquid nitrogen, and data was collected at the Australian Synchrotron (MX2 beamline) and processed using XDS [209]. The structure was solved by molecular replacement using PHASER [210] using a search model created by removing the variable loops from a IGG2 FAB (PDB ID: 5MYK) [211]. The final model was built in Coot [212] and refined with Phenix [213] to a resolution of 2.00 Å. Data collection and refinement statistics are summarized in Table 6.1. The R_{work} and R_{free} after the final refinement was 0.2507 and 0.2651 respectively. The coordinates have not yet been deposited in the Protein Data Bank. Figures were prepared using Pymol.

8.6 - Materials

Figure 8. 5: List of antibodies

Reagent	Source	Identifier
Monoclonal ANTI-FLAG® M2 antibody	Sigma	F1804
Lipopolysaccharide (LPS) Monoclonal Antibody	Biomatik	CAU29364
Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP	ThermoFisher	62-6520
Anti-GAPDH antibody	Sigma	G9545
Goat Anti-Rabbit IgG H&L (HRP)	Abcam	ab6721

Figure 8. 6: List of enzymes

Reagent	Source	Identifier
Quick Ligase	New England Biolabs (NEB)	M2200S
NotI-HF	New England Biolabs (NEB)	R3189S
BamHI-HF	New England Biolabs (NEB)	R3136S
EcoRI-HF	New England Biolabs (NEB)	R3101S
XhoI	New England Biolabs (NEB)	R0146S
SalI-HF	New England Biolabs (NEB)	R3138S
AvRII-HF	New England Biolabs (NEB)	R0174S
Trypsin Gold, Mass Spectrometry Grade	Promega	V5280
Lyticase	Sigma	L2524
Clostripain (Endoproteinase-Arg-C)	Worthington	LS001643
Recombinant shrimp alkaline phosphatase (rSAP)	New England Biolabs (NEB)	M0371S

Figure 8. 7: List of reagents

Reagent	Source	Identifier
Phorbol 12-myristate 13-acetate (PMA)	Sigma	P1585
FuGENE® HD Transfection Reagent	Promega	E2311
Polybrene	Sigma	107689
Propidium iodide	Sigma	81845
Bovine Serum Albumin	Sigma	A2058
OPD (o-phenylenediamine dihydrochloride)	ThermoFisher	34005
cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail	Roche	11836170001
Immobilon Western Chemiluminescent HRP Substrate	Merck	WBKLS0500
Sep-Pak C18 Plus Short Cartridge	Waters	WAT020515
Nitrocellulose Membrane, 0.45 µm	ThermoFisher	88018
Dulbecco's phosphate-buffered saline (DPBS)	ThermoFisher	14190250
GeneRuler 1 kb Plus DNA Ladder	ThermoFisher	SM1331
Yeast Extract	Sigma	92144
Peptone from casein (Tryptone)	Merck	107213
Yeast Nitrogen Base Without Amino Acids	Sigma	Y0626
SYPRO™ Orange Protein Gel Stain	ThermoFisher	S6650
Ribonucleic acid from torula yeast	Sigma	R6625

Sodium meta-periodate	ThermoFisher	20504
Amino-oxy-biotin	Biotium	90113
Streptavidin−Agarose	Sigma	S1638
Dolichol-Phosphate	Avanti	Custom
Guanosine 5′-diphospho-D-mannose sodium salt	Sigma	G5131
Tritiated GDP-Mannose	ARC	ART 0440
Emulsifier-Safe	PerkinElmer	6013389
WAKW, WSEW, WSEA, WSEF, wSEW, WSEw, wSEw, WGGW synthetic peptides	Mimotopes	Custom

Figure 8. 8: Composition of media

Media	**Composition**
RPMI	(Thermofisher, A1049101) 1x Penicillin-Streptomycin solution
Dulbecco's Modified Eagle's (DME) medium	DMEM powder (Gibco/Thermo, Cat#31600083) 40 mM $NaHCO_3$ 1x Penicillin-Streptomycin solution (Sigma, Cat#P0781)
FACs buffer (KDS-BSS, 2-3% FBS)	150 mM NaCl 3.7 mM KCl 2.5 mM $CaCl_2$ 1.2 mM $MgSO_4$ 1.2 mM KH_2PO_4 0.8 mM K_2HPO_4 7.4 mM HEPES NaOH
Phosphate buffered saline (PBS) pH 7.3	20 mM Na_2HPO_4 4.5 mM NaH_2PO_4 150 mM NaCl
10X YNB (13.4% yeast nitrogen base with ammonium sulfate without amino acids)	134 g of yeast nitrogen base (YNB) Make to 1L with MilliQ water and filter.
500X B (0.02% Biotin)	20 mg biotin in 100 mL of water and filter sterilize.
10X D (20% Dextrose)	200 g of D-glucose in 1,000 mL of water
10X M (5% methanol)	5 mL of methanol in 95 mL of water
10X GY (10% glycerol)	100 mL of glycerol in 900 mL of water
Yeast Extract Peptone Dextrose Medium (YPD)	1% yeast extract 2% peptone 2% dextrose (glucose)
YPD-zeocin plates	1% yeast extract 2% peptone 2% dextrose (glucose) 2% agar 100ug/ml zeocin
Minimal Dextrose plates (MD)	1.34% YNB 4 × 10-5 % biotin 2% dextrose

	2% agar
Minimal Methanol plates (MM)	1.34% YNB 4 × 10-5 % biotin 0.5% methanol 2% agar
Buffered Glycerol-complex Medium (BMGY)	1% yeast extract 2% peptone 100 mM potassium phosphate, pH 6.0 1.34% YNB 4 × 10-5 % biotin 1% glycerol
Buffered Methanol-complex Medium (BMMY)	1% yeast extract 2% peptone 100 mM potassium phosphate, pH 6.0 1.34% YNB 4 × 10-5 % biotin 0.5% methanol

Figure 8. 9: List of buffer compositions

Buffer	**Composition**
1% Triton X-100 lysis buffer	Triton X-100 20 mM Tris-HCl, pH 7.5 150 mM NaCl 2 mM EDTA 1 Protease Inhibitor Cocktail tablet (Roche)
RIPA lysis buffer	25 mM Tris-HCl, pH 7.5 150 mM NaCl 2 mM EDTA 1% v/v Triton X-100 1% w/v sodium dodecyl sulfate 0.1% w/v SDS 1 Protease Inhibitor Cocktail tablet (Roche)
4x Sample reducing buffer	250 mM Tris–HCl, pH 6.8 40% glycerol 8% SDS 0.2% bromophenol blue 200 mM DTT
4x Sample non-reducing buffer	250 mM Tris–HCl, pH 6.8 40% glycerol 8% SDS 0.2% bromophenol blue
Tris-buffered saline (TBS)	20 mM Tris-HCl 150 mM NaCl pH 7.5
IMAC dialysis buffer	50 mM Tris-HCl 150 mM NaCl pH 7.5
IMAC wash buffer	50 mM Tris-HCl 150 mM NaCl 5 mM Imidazole pH 7.5

IMAC elution buffer	50 mM Tris-HCl 150 mM NaCl 500 mM Imidazole pH 7.5
MonoQ binding buffer	50 mM Tris-HCl pH 7.5
MonoQ elution buffer	50 mM Tris-HCl 500 mM NaCl pH 7.5
Yeast cracking buffer	30 mM Tris-HC1, pH 7.5 50 mM KC1 5 mM $MgC1_2$ 1 M sorbitol 10 mM DTT 20% v/v glycerol

9 - References

1. Spiro, R.G., *Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds.* Glycobiology, 2002. **12**(4): p. 43R-56R.
2. Mann, M. and O.N. Jensen, *Proteomic analysis of post-translational modifications.* Nat Biotechnol, 2003. **21**(3): p. 255-61.
3. Apweiler, R., H. Hermjakob, and N. Sharon, *On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database.* Biochim Biophys Acta, 1999. **1473**(1): p. 4-8.
4. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function.* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 448-62.
5. Ohtsubo, K. and J.D. Marth, *Glycosylation in cellular mechanisms of health and disease.* Cell, 2006. **126**(5): p. 855-67.
6. Neuberger, A., *Carbohydrates in protein: The carbohydrate component of crystalline egg albumin.* Biochem J, 1938. **32**(9): p. 1435-51.
7. Andersson, L.C., M. Jokinen, and C.G. Gahmberg, *Induction of erythroid differentiation in the human leukaemia cell line K562.* Nature, 1979. **278**(5702): p. 364-5.
8. Hofsteenge, J., et al., *New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us.* Biochemistry, 1994. **33**(46): p. 13524-30.
9. de Beer, T., et al., *The hexopyranosyl residue that is C-glycosidically linked to the side chain of tryptophan-7 in human RNase Us is alpha-mannopyranose.* Biochemistry, 1995. **34**(37): p. 11785-9.
10. Furmanek, A., et al., *The WSAWS motif is C-hexosylated in a soluble form of the erythropoietin receptor.* Biochemistry, 2003. **42**(28): p. 8452-8.
11. Hofsteenge, J., et al., *The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues.* J Biol Chem, 1999. **274**(46): p. 32786-94.
12. Hamming, O.J., et al., *Crystal structure of interleukin-21 receptor (IL-21R) bound to IL-21 reveals that sugar chain interacting with WSXWS motif is integral part of IL-21R.* J Biol Chem, 2012. **287**(12): p. 9454-60.
13. Krieg, J., et al., *Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp.* Mol Biol Cell, 1998. **9**(2): p. 301-9.
14. Doucey, M.A., et al., *Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosphate-mannose as a precursor.* Mol Biol Cell, 1998. **9**(2): p. 291-300.
15. Sharma, V., M. Ichikawa, and H.H. Freeze, *Mannose metabolism: more than meets the eye.* Biochem Biophys Res Commun, 2014. **453**(2): p. 220-8.
16. Hirschberg, C.B. and M.D. Snider, *Topography of glycosylation in the rough endoplasmic reticulum and Golgi apparatus.* Annu Rev Biochem, 1987. **56**: p. 63-87.
17. Olsen, J.G. and B.B. Kragelund, *Who climbs the tryptophan ladder? On the structure and function of the WSXWS motif in cytokine receptors and thrombospondin repeats.* Cytokine Growth Factor Rev, 2014. **25**(3): p. 337-41.
18. Gallivan, J.P. and D.A. Dougherty, *Cation-pi interactions in structural biology.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9459-64.
19. Flocco, M.M. and S.L. Mowbray, *Planar stacking interactions of arginine and aromatic side-chains in proteins.* J Mol Biol, 1994. **235**(2): p. 709-17.
20. Craven, T.W., et al., *A Miniature Protein Stabilized by a Cation-pi Interaction Network.* J Am Chem Soc, 2016. **138**(5): p. 1543-50.

21. Siupka, P., et al., *A conserved sugar bridge connected to the WSXWS motif has an important role for transport of IL-21R to the plasma membrane.* Genes Immun, 2015. **16**(6): p. 405-13.
22. Yu, H., et al., *Specificities of heparin-binding sites from the amino-terminus and type 1 repeats of thrombospondin-1.* Arch Biochem Biophys, 2000. **374**(1): p. 13-23.
23. Julenius, K., *NetCGlyc 1.0: prediction of mammalian C-mannosylation sites.* Glycobiology, 2007. **17**(8): p. 868-76.
24. de Castro, E., et al., *ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W362-5.
25. Emanuelsson, O., et al., *Locating proteins in the cell using TargetP, SignalP and related tools.* Nat Protoc, 2007. **2**(4): p. 953-71.
26. Thomas, P.D., et al., *PANTHER: a library of protein families and subfamilies indexed by function.* Genome Res, 2003. **13**(9): p. 2129-41.
27. Mi, H., A. Muruganujan, and P.D. Thomas, *PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees.* Nucleic Acids Res, 2013. **41**(Database issue): p. D377-86.
28. Nishimura, R., et al., *The role of Smads in BMP signaling.* Front Biosci, 2003. **8**: p. s275-84.
29. Gilman, A.G., *G proteins: transducers of receptor-generated signals.* Annu Rev Biochem, 1987. **56**: p. 615-49.
30. Das, S., et al., *Brain angiogenesis inhibitor 1 is expressed by gastric phagocytes during infection with Helicobacter pylori and mediates the recognition and engulfment of human apoptotic gastric epithelial cells.* Faseb j, 2014. **28**(5): p. 2214-24.
31. Malnic, B., P.A. Godfrey, and L.B. Buck, *The human olfactory receptor gene family.* Proc Natl Acad Sci U S A, 2004. **101**(8): p. 2584-9.
32. Bufe, B., et al., *The human TAS2R16 receptor mediates bitter taste in response to beta-glucopyranosides.* Nat Genet, 2002. **32**(3): p. 397-401.
33. Sudhof, T.C., et al., *The LDL receptor gene: a mosaic of exons shared with different proteins.* Science, 1985. **228**(4701): p. 815-22.
34. Samatov, T.R., D. Wicklein, and A.G. Tonevitsky, *L1CAM: Cell adhesion and more.* Prog Histochem Cytochem, 2016. **51**(2): p. 25-32.
35. Moore, S.W., M. Tessier-Lavigne, and T.E. Kennedy, *Netrins and their receptors.* Adv Exp Med Biol, 2007. **621**: p. 17-31.
36. Dubois, C.M., et al., *Processing of transforming growth factor beta 1 precursor by human furin convertase.* J Biol Chem, 1995. **270**(18): p. 10618-24.
37. Minic, J., et al., *Butyrylcholinesterase and acetylcholinesterase activity and quantal transmitter release at normal and acetylcholinesterase knockout mouse neuromuscular junctions.* Br J Pharmacol, 2003. **138**(1): p. 177-87.
38. Zisman, L.S., *ACE and ACE2: a tale of two enzymes.* Eur Heart J, 2005. **26**(4): p. 322-4.
39. Taron, B.W., et al., *Human Smp3p adds a fourth mannose to yeast and human glycosylphosphatidylinositol precursors in vivo.* J Biol Chem, 2004. **279**(34): p. 36083-92.
40. Takeuchi, H., et al., *Rumi functions as both a protein O-glucosyltransferase and a protein O-xylosyltransferase.* Proc Natl Acad Sci U S A, 2011. **108**(40): p. 16600-5.
41. Eriksson, U. and K. Alitalo, *Structure, expression and receptor-binding properties of novel vascular endothelial growth factors.* Curr Top Microbiol Immunol, 1999. **237**: p. 41-57.

42. Barlesi, F., et al., *Randomized phase III trial of maintenance bevacizumab with or without pemetrexed after first-line induction with bevacizumab, cisplatin, and pemetrexed in advanced nonsquamous non-small-cell lung cancer: AVAPERL (MO22089).* J Clin Oncol, 2013. **31**(24): p. 3004-11.
43. Meno, C., et al., *lefty-1 is required for left-right determination as a regulator of lefty-2 and nodal.* Cell, 1998. **94**(3): p. 287-97.
44. Raslan, A.A. and J.K. Yoon, *R-spondins: Multi-mode WNT signaling regulators in adult stem cells.* Int J Biochem Cell Biol, 2019. **106**: p. 26-34.
45. Hall, N.G., et al., *ADAMTSL-3/punctin-2, a novel glycoprotein in extracellular matrix related to the ADAMTS family of metalloproteases.* Matrix Biol, 2003. **22**(6): p. 501-10.
46. Abdelmagid, S.M., et al., *Osteoactivin, an anabolic factor that regulates osteoblast differentiation and function.* Exp Cell Res, 2008. **314**(13): p. 2334-51.
47. Patil, R.V., et al., *Expression of aquaporins in the rat ocular tissue.* Exp Eye Res, 1997. **64**(2): p. 203-9.
48. Liongue, C. and A.C. Ward, *Evolution of Class I cytokine receptors.* BMC Evol Biol, 2007. **7**: p. 120.
49. Wang, X., et al., *Structural biology of shared cytokine receptors.* Annu Rev Immunol, 2009. **27**: p. 29-60.
50. Bagley, C.J., et al., *The structural and functional basis of cytokine receptor activation: lessons from the common beta subunit of the granulocyte-macrophage colony-stimulating factor, interleukin-3 (IL-3), and IL-5 receptors.* Blood, 1997. **89**(5): p. 1471-82.
51. Bazan, J.F., *Structural design and molecular evolution of a cytokine receptor superfamily.* Proc Natl Acad Sci U S A, 1990. **87**(18): p. 6934-8.
52. Taga, T. and T. Kishimoto, *Gp130 and the interleukin-6 family of cytokines.* Annu Rev Immunol, 1997. **15**: p. 797-819.
53. Kishimoto, T., et al., *Interleukin-6 family of cytokines and gp130.* Blood, 1995. **86**(4): p. 1243-54.
54. Sato, N. and A. Miyajima, *Multimeric cytokine receptors: common versus specific functions.* Curr Opin Cell Biol, 1994. **6**(2): p. 174-9.
55. Coquet, J.M., et al., *IL-21 is produced by NKT cells and modulates NKT cell activation and cytokine production.* J Immunol, 2007. **178**(5): p. 2827-34.
56. Kaushansky, K., et al., *Promotion of megakaryocyte progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin.* Nature, 1994. **369**(6481): p. 568-71.
57. Sasazawa, Y., et al., *C-mannosylation of thrombopoietin receptor (c-Mpl) regulates thrombopoietin-dependent JAK-STAT signaling.* Biochemical and Biophysical Research Communications, 2015(1-2): p. 262.
58. Watowich, S.S., *The erythropoietin receptor: molecular structure and hematopoietic signaling pathways.* J Investig Med, 2011. **59**(7): p. 1067-72.
59. Hilton, D.J., et al., *Saturation mutagenesis of the WSXWS motif of the erythropoietin receptor.* J Biol Chem, 1996. **271**(9): p. 4699-708.
60. Kobayashi, M., et al., *Identification and purification of natural killer cell stimulatory factor (NKSF), a cytokine with multiple biologic effects on human lymphocytes.* J Exp Med, 1989. **170**(3): p. 827-45.
61. Doucey, M.A., et al., *Recombinant human interleukin-12 is the second example of a C-mannosylated protein.* Glycobiology, 1999. **9**(5): p. 435-41.

62. Adams, J.C. and R.P. Tucker, *The thrombospondin type 1 repeat (TSR) superfamily: diverse proteins with related roles in neuronal development.* Dev Dyn, 2000. **218**(2): p. 280-99.
63. Wang, L.W., et al., *Post-translational modification of thrombospondin type-1 repeats in ADAMTS-like 1/punctin-1 by C-mannosylation of tryptophan.* J Biol Chem, 2009. **284**(44): p. 30004-15.
64. Hofsteenge, J., et al., *C-mannosylation and O-fucosylation of the thrombospondin type 1 module.* J Biol Chem, 2001. **276**(9): p. 6485-98.
65. Hartmann, S. and J. Hofsteenge, *Properdin, the positive regulator of complement, is highly C-mannosylated.* J Biol Chem, 2000. **275**(37): p. 28569-74.
66. Lovelace, L.L., et al., *Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement.* J Biol Chem, 2011. **286**(20): p. 17585-92.
67. Leung-Hagesteijn, C., et al., *UNC-5, a transmembrane protein with immunoglobulin and thrombospondin type 1 domains, guides cell and pioneer axon migrations in C. elegans.* Cell, 1992. **71**(2): p. 289-99.
68. Buettner, F.F., et al., *C. elegans DPY-19 is a C-mannosyltransferase glycosylating thrombospondin repeats.* Mol Cell, 2013. **50**(2): p. 295-302.
69. Li, Y., et al., *Structure of the F-spondin domain of mindin, an integrin ligand and pattern recognition molecule.* EMBO J, 2009. **28**(3): p. 286-97.
70. Gonzalez de Peredo, A., et al., *C-mannosylation and o-fucosylation of thrombospondin type 1 repeats.* Mol Cell Proteomics, 2002. **1**(1): p. 11-8.
71. Niwa, Y., et al., *Identification of DPY19L3 as the C-mannosyltransferase of R-spondin1 in human cells.* Mol Biol Cell, 2016.
72. Morishita, S., et al., *Dpy-19 like 3-mediated C-mannosylation and expression levels of RPE-spondin in human tumor cell lines.* Oncol Lett, 2017. **14**(2): p. 2537-2544.
73. Fujiwara, M., et al., *C-mannosylation of R-spondin3 regulates its secretion and activity of Wnt/beta-catenin signaling in cells.* FEBS Lett, 2016. **590**(16): p. 2639-49.
74. Shcherbakova, A., et al., *Distinct C-mannosylation of netrin receptor thrombospondin type 1 repeats by mammalian DPY19L1 and DPY19L3.* Proc Natl Acad Sci U S A, 2017. **114**(10): p. 2574-2579.
75. Klar, A., M. Baldassare, and T.M. Jessell, *F-spondin: a gene expressed at high levels in the floor plate encodes a secreted protein that promotes neural cell adhesion and neurite extension.* Cell, 1992. **69**(1): p. 95-110.
76. Higashijima, S., et al., *Mindin/F-spondin family: novel ECM proteins expressed in the zebrafish embryonic axis.* Dev Biol, 1997. **192**(2): p. 211-27.
77. Feinstein, Y., et al., *F-spondin and mindin: two structurally and functionally related genes expressed in the hippocampus that promote outgrowth of embryonic hippocampal neurons.* Development, 1999. **126**(16): p. 3637-48.
78. Li, H., et al., *Efficient dendritic cell priming of T lymphocytes depends on the extracellular matrix protein mindin.* Embo j, 2006. **25**(17): p. 4097-107.
79. Jia, W., H. Li, and Y.W. He, *The extracellular matrix protein mindin serves as an integrin ligand and is critical for inflammatory cell recruitment.* Blood, 2005. **106**(12): p. 3854-9.
80. Jia, W., H. Li, and Y.W. He, *Pattern recognition molecule mindin promotes intranasal clearance of influenza viruses.* J Immunol, 2008. **180**(9): p. 6255-61.
81. He, Y.W., et al., *The extracellular matrix protein mindin is a pattern-recognition molecule for microbial pathogens.* Nat Immunol, 2004. **5**(1): p. 88-97.
82. Hourcade, D.E., *The role of properdin in the assembly of the alternative pathway C3 convertases of complement.* J Biol Chem, 2006. **281**(4): p. 2128-32.

83. Noris, M. and G. Remuzzi, *Overview of complement activation and regulation.* Semin Nephrol, 2013. **33**(6): p. 479-92.
84. Holers, V.M., *Complement and its receptors: new insights into human disease.* Annu Rev Immunol, 2014. **32**: p. 433-59.
85. Upshaw, J.D., Jr., *Congenital deficiency of a factor in normal plasma that reverses microangiopathic hemolysis and thrombocytopenia.* N Engl J Med, 1978. **298**(24): p. 1350-2.
86. Zheng, X., et al., *Structure of von Willebrand factor-cleaving protease (ADAMTS13), a metalloprotease involved in thrombotic thrombocytopenic purpura.* J Biol Chem, 2001. **276**(44): p. 41059-63.
87. Sorvillo, N., et al., *Identification of N-linked glycosylation and putative O-fucosylation, C-mannosylation sites in plasma derived ADAMTS13.* J Thromb Haemost, 2014. **12**(5): p. 670-9.
88. Ling, J., et al., *The WXXW motif in the TSR1 of ADAMTS13 is important for its secretion and proteolytic activity.* Thromb Res, 2013. **131**(6): p. 529-34.
89. Rogers, D.F., *Airway mucus hypersecretion in asthma: an undervalued pathology?* Curr Opin Pharmacol, 2004. **4**(3): p. 241-50.
90. Kirkham, S., et al., *Heterogeneity of airways mucus: variations in the amounts and glycoforms of the major oligomeric mucins MUC5AC and MUC5B.* Biochem J, 2002. **361**(Pt 3): p. 537-46.
91. Groneberg, D.A., et al., *Expression of MUC5AC and MUC5B mucins in normal and cystic fibrosis lung.* Respir Med, 2002. **96**(2): p. 81-6.
92. Perez-Vilar, J., S.H. Randell, and R.C. Boucher, *C-Mannosylation of MUC5AC and MUC5B Cys subdomains.* Glycobiology, 2004. **14**(4): p. 325-37.
93. Gouyer, V., et al., *Non-C-mannosylable mucin CYS domains hindered proper folding and secretion of mucin.* Biochem Biophys Res Commun, 2018. **506**(4): p. 812-818.
94. Toole, B.P., *Hyaluronan is not just a goo!* J Clin Invest, 2000. **106**(3): p. 335-6.
95. Chao, K.L., L. Muthukumar, and O. Herzberg, *Structure of human hyaluronidase-1, a hyaluronan hydrolyzing enzyme involved in tumor growth and angiogenesis.* Biochemistry, 2007. **46**(23): p. 6911-20.
96. Goto, Y., et al., *C-mannosylation of human hyaluronidase 1: possible roles for secretion and enzymatic activity.* Int J Oncol, 2014. **45**(1): p. 344-50.
97. Okamoto, S., et al., *Regulation of secretion and enzymatic activity of lipoprotein lipase by C-mannosylation.* Biochem Biophys Res Commun, 2017. **486**(2): p. 558-563.
98. Braun, J.E. and D.L. Severson, *Regulation of the synthesis, processing and translocation of lipoprotein lipase.* Biochem J, 1992. **287 (Pt 2)**: p. 337-47.
99. Mukhopadhyay, G., et al., *A novel role for myelin-associated glycoprotein as an inhibitor of axonal regeneration.* Neuron, 1994. **13**(3): p. 757-67.
100. Trapp, B.D., R.H. Quarles, and J.W. Griffin, *Myelin-associated glycoprotein and myelinating Schwann cell-axon interaction in chronic B,B'-iminodipropionitrile neuropathy.* J Cell Biol, 1984. **98**(4): p. 1272-8.
101. Yang, L.J., et al., *Gangliosides are neuronal ligands for myelin-associated glycoprotein.* Proc Natl Acad Sci U S A, 1996. **93**(2): p. 814-8.
102. Pronker, M.F., et al., *Structural basis of myelin-associated glycoprotein adhesion and signalling.* Nat Commun, 2016. **7**: p. 13584.
103. Wyatt, G.R. and G.F. Kale, *The chemistry of insect hemolymph. II. Trehalose and other carbohydrates.* J Gen Physiol, 1957. **40**(6): p. 833-47.
104. Munte, C.E., et al., *C-mannosylation in the hypertrehalosaemic hormone from the stick insect Carausius morosus.* Febs j, 2008. **275**(6): p. 1163-73.

105. Gade, G., et al., *A tryptophan-substituted member of the AKH/RPCH family isolated from a stick insect corpus cardiacum.* Biochem Biophys Res Commun, 1992. **189**(3): p. 1303-9.
106. Zhao, H., et al., *Glycosylated hydroxytryptophan in a mussel adhesive protein from Perna viridis.* J Biol Chem, 2009. **284**(35): p. 23344-52.
107. Petrone, L., et al., *Mussel adhesion is dictated by time-regulated secretion and molecular conformation of mussel adhesive proteins.* Nat Commun, 2015. **6**: p. 8737.
108. Hwang, D.S., et al., *Adhesion mechanism in a DOPA-deficient foot protein from green mussels().* Soft Matter, 2012. **8**(20): p. 5640-5648.
109. Ervin, L.A., et al., *Phosphorylation and glycosylation of bovine lens MP20.* Invest Ophthalmol Vis Sci, 2005. **46**(2): p. 627-35.
110. Gonen, T., et al., *MP20, the second most abundant lens membrane protein and member of the tetraspanin superfamily, joins the list of ligands of galectin-3.* BMC Cell Biol, 2001. **2**: p. 17.
111. Falzarano, D., et al., *Ebola sGP--the first viral glycoprotein shown to be C-mannosylated.* Virology, 2007. **368**(1): p. 83-90.
112. Pfeffer, S., et al., *Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon.* Nat Commun, 2014. **5**: p. 3072.
113. Honigberg, L. and C. Kenyon, *Establishment of left/right asymmetry in neuroblast migration by UNC-40/DCC, UNC-73/Trio and DPY-19 proteins in C. elegans.* Development, 2000. **127**(21): p. 4655-68.
114. Carson, A.R., J. Cheung, and S.W. Scherer, *Duplication and relocation of the functional DPY19L2 gene within low copy repeats.* BMC Genomics, 2006. **7**: p. 45.
115. Walsh, B., *Population-genetic models of the fates of duplicate genes.* Genetica, 2003. **118**(2-3): p. 279-94.
116. Koscinski, I., et al., *DPY19L2 deletion as a major cause of globozoospermia.* Am J Hum Genet, 2011. **88**(3): p. 344-50.
117. Harbuz, R., et al., *A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation.* Am J Hum Genet, 2011. **88**(3): p. 351-61.
118. Ghedir, H., et al., *Identification of a new DPY19L2 mutation and a better definition of DPY19L2 deletion breakpoints leading to globozoospermia.* Mol Hum Reprod, 2016. **22**(1): p. 35-45.
119. Zhu, F., et al., *DPY19L2 gene mutations are a major cause of globozoospermia: identification of three novel point mutations.* Mol Hum Reprod, 2013. **19**(6): p. 395-404.
120. Pierre, V., et al., *Absence of Dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus.* Development, 2012. **139**(16): p. 2955-65.
121. Sidik, S.M., et al., *A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes.* Cell, 2016. **166**(6): p. 1423-1435.e12.
122. Hoppe, C.M., et al., *Apicomplexan C-Mannosyltransferases Modify Thrombospondin Type I-containing Adhesins of the TRAP Family.* Glycobiology, 2018. **28**(5): p. 333-343.
123. Khurana, S., et al., *Protein O-fucosyltransferase 2-mediated O-glycosylation of the adhesin MIC2 is dispensable for Toxoplasma gondii tachyzoite infection.* J Biol Chem, 2018.
124. Swearingen, K.E., et al., *Interrogating the Plasmodium Sporozoite Surface: Identification of Surface-Exposed Proteins and Demonstration of Glycosylation on*

CSP and TRAP by Mass Spectrometry-Based Proteomics. PLoS Pathog, 2016. **12**(4): p. e1005606.
125. Tewari, R., et al., *Function of region I and II adhesive motifs of Plasmodium falciparum circumsporozoite protein in sporozoite motility and infectivity.* J Biol Chem, 2002. **277**(49): p. 47613-8.
126. Coppi, A., et al., *The malaria circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host.* J Exp Med, 2011. **208**(2): p. 341-56.
127. Labaied, M., N. Camargo, and S.H. Kappe, *Depletion of the Plasmodium berghei thrombospondin-related sporozoite protein reveals a role in host cell entry by sporozoites.* Mol Biochem Parasitol, 2007. **153**(2): p. 158-66.
128. Zanetta, J.P., et al., *Quantitative gas chromatography/mass spectrometry determination of C-mannosylation of tryptophan residues in glycoproteins.* Anal Biochem, 2004. **329**(2): p. 199-206.
129. Bornholdt, Z.A., et al., *Isolation of potent neutralizing antibodies from a survivor of the 2014 Ebola virus outbreak.* Science, 2016. **351**(6277): p. 1078-83.
130. Bauza, K., et al., *Tailoring a Combination Preerythrocytic Malaria Vaccine.* Infect Immun, 2015. **84**(3): p. 622-34.
131. Nakayama, K., et al., *OCH1 encodes a novel membrane bound mannosyltransferase: outer chain elongation of asparagine-linked oligosaccharides.* Embo j, 1992. **11**(7): p. 2511-9.
132. Callewaert, N., et al., *Use of HDEL-tagged Trichoderma reesei mannosyl oligosaccharide 1,2-alpha-D-mannosidase for N-glycan engineering in Pichia pastoris.* FEBS Lett, 2001. **503**(2-3): p. 173-8.
133. Vervecken, W., et al., *In vivo synthesis of mammalian-like, hybrid-type N-glycans in Pichia pastoris.* Appl Environ Microbiol, 2004. **70**(5): p. 2639-46.
134. Hamilton, S.R., et al., *Production of complex human glycoproteins in yeast.* Science, 2003. **301**(5637): p. 1244-6.
135. Choi, B.K., et al., *Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast Pichia pastoris.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5022-7.
136. Jacobs, P.P., et al., *Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology.* Nat Protoc, 2009. **4**(1): p. 58-70.
137. Hamilton, S.R., et al., *Humanization of yeast to produce complex terminally sialylated glycoproteins.* Science, 2006. **313**(5792): p. 1441-3.
138. Li, H., et al., *Optimization of humanized IgGs in glycoengineered Pichia pastoris.* Nat Biotechnol, 2006. **24**(2): p. 210-5.
139. Nett, J.H., et al., *Characterization of the Pichia pastoris protein-O-mannosyltransferase gene family.* PLoS One, 2013. **8**(7): p. e68325.
140. Gomathinayagam, S. and S.R. Hamilton, *In vitro enzymatic treatment to remove O-linked mannose from intact glycoproteins.* Appl Microbiol Biotechnol, 2014. **98**(6): p. 2545-54.
141. Hamilton, S.R., et al., *Production of sialylated O-linked glycans in Pichia pastoris.* Glycobiology, 2013. **23**(10): p. 1192-203.
142. Amano, K., et al., *Engineering of mucin-type human glycoproteins in yeast cells.* Proc Natl Acad Sci U S A, 2008. **105**(9): p. 3232-7.
143. Orlean, P., C. Albright, and P.W. Robbins, *Cloning and sequencing of the yeast gene for dolichol phosphate mannose synthase, an essential protein.* J Biol Chem, 1988. **263**(33): p. 17499-507.

144. Rath, A., et al., *Detergent binding explains anomalous SDS-PAGE migration of membrane proteins.* Proc Natl Acad Sci U S A, 2009. **106**(6): p. 1760-5.
145. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast.* Nature, 2003. **425**(6959): p. 737-41.
146. Cereghino, J.L. and J.M. Cregg, *Heterologous protein expression in the methylotrophic yeast Pichia pastoris.* FEMS Microbiol Rev, 2000. **24**(1): p. 45-66.
147. Ellis, S.B., et al., *Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris.* Mol Cell Biol, 1985. **5**(5): p. 1111-21.
148. Hopkins, D., et al., *Elimination of diaminopeptidase activity in Pichia pastoris for therapeutic protein production.* Appl Microbiol Biotechnol, 2014. **98**(6): p. 2573-83.
149. Yang, S., et al., *Enhanced production of recombinant secretory proteins in Pichia pastoris by optimizing Kex2 P1' site.* PLoS One, 2013. **8**(9): p. e75347.
150. De Wachter, C., L. Van Landuyt, and N. Callewaert, *Engineering of Yeast Glycoprotein Expression.* Adv Biochem Eng Biotechnol, 2018.
151. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. **2**(9): p. 2212-21.
152. Yang, D., et al., *Eosinophil-derived neurotoxin (EDN), an antimicrobial protein with chemotactic activities for dendritic cells.* Blood, 2003. **102**(9): p. 3396-403.
153. Yang, D., et al., *Eosinophil-derived neurotoxin acts as an alarmin to activate the TLR2-MyD88 signal pathway in dendritic cells and enhances Th2 immune responses.* J Exp Med, 2008. **205**(1): p. 79-90.
154. Domachowske, J.B., et al., *Recombinant human eosinophil-derived neurotoxin/RNase 2 functions as an effective antiviral agent against respiratory syncytial virus.* J Infect Dis, 1998. **177**(6): p. 1458-64.
155. Lee-Huang, S., et al., *Lysozyme and RNases as anti-HIV components in beta-core preparations of human chorionic gonadotropin.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2678-81.
156. Yang, D., et al., *Human ribonuclease A superfamily members, eosinophil-derived neurotoxin and pancreatic ribonuclease, induce dendritic cell maturation and activation.* J Immunol, 2004. **173**(10): p. 6134-42.
157. Tripathy, D.R., A.K. Dinda, and S. Dasgupta, *A simple assay for the ribonuclease activity of ribonucleases in the presence of ethidium bromide.* Anal Biochem, 2013. **437**(2): p. 126-9.
158. Martinez-Sernandez, V., et al., *Usefulness of ELISA Methods for Assessing LPS Interactions with Proteins and Peptides.* PLoS One, 2016. **11**(6): p. e0156530.
159. Brocke-Heidrich, K., et al., *Interleukin-6-dependent gene expression profiles in multiple myeloma INA-6 cells reveal a Bcl-2 family-independent survival pathway closely associated with Stat3 activation.* Blood, 2004. **103**(1): p. 242-51.
160. Barata, J.T., et al., *Interleukin-7 promotes survival and cell cycle progression of T-cell acute lymphoblastic leukemia cells by down-regulating the cyclin-dependent kinase inhibitor p27(kip1).* Blood, 2001. **98**(5): p. 1524-31.
161. Young, D.A., et al., *Blockade of the interleukin-21/interleukin-21 receptor pathway ameliorates disease in animal models of rheumatoid arthritis.* Arthritis Rheum, 2007. **56**(4): p. 1152-63.
162. Wetzel-Smith, M.K., et al., *A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death.* Nat Med, 2014. **20**(12): p. 1452-7.
163. Park, D., et al., *BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module.* Nature, 2007. **450**(7168): p. 430-4.

164. Niwa, Y., et al., *Topological analysis of DPY19L3, a human C-mannosyltransferase.* Febs j, 2018. **285**(6): p. 1162-1174.
165. Ihara, Y., et al., *Increased expression of protein C-mannosylation in the aortic vessels of diabetic Zucker rats.* Glycobiology, 2005. **15**(4): p. 383-92.
166. Nishikawa, T., et al., *Total Synthesis of α-C-Mannosyltryptophan, a Naturally Occurring C-Glycosyl Amino Acid.* Synlett, 2001. **2001**(Special Issue): p. 0945-0947.
167. Manabe, S., Y. Marui, and Y. Ito, *Total synthesis of mannosyl tryptophan and its derivatives.* Chemistry, 2003. **9**(6): p. 1435-47.
168. Nishikawa, T., et al., *Stereocontrolled syntheses of alpha-C-mannosyltryptophan and its analogues.* Org Biomol Chem, 2005. **3**(4): p. 687-700.
169. Mitchell, W.M., *Cleavage at arginine residues by clostripain.* Methods Enzymol, 1977. **47**: p. 165-70.
170. Markl, J., et al., *Marine tumor vaccine carriers: structure of the molluscan hemocyanins KLH and htH.* J Cancer Res Clin Oncol, 2001. **127 Suppl 2**: p. R3-9.
171. Gatsogiannis, C. and J. Markl, *Keyhole limpet hemocyanin: 9-A CryoEM structure and molecular model of the KLH1 didecamer reveal the interfaces and intricate topology of the 160 functional units.* J Mol Biol, 2009. **385**(3): p. 963-83.
172. Kohler, G. and C. Milstein, *Continuous cultures of fused cells secreting antibody of predefined specificity.* Nature, 1975. **256**(5517): p. 495-7.
173. Margulies, D.H., W.M. Kuehl, and M.D. Scharff, *Somatic cell hybridization of mouse myeloma cells.* Cell, 1976. **8**(3): p. 405-15.
174. Kohler, G., S.C. Howe, and C. Milstein, *Fusion between immunoglobulin-secreting and nonsecreting myeloma cell lines.* Eur J Immunol, 1976. **6**(4): p. 292-5.
175. Mou, Y., et al., *Engineering Improved Antiphosphotyrosine Antibodies Based on an Immunoconvergent Binding Motif.* J Am Chem Soc, 2018. **140**(48): p. 16615-16624.
176. Patel, D.S., et al., *Conformational properties of alpha- or beta-(1-->6)-linked oligosaccharides: Hamiltonian replica exchange MD simulations and NMR experiments.* J Phys Chem B, 2014. **118**(11): p. 2851-71.
177. Nakajima, T., et al., *Soluble interleukin-6 receptor is released from receptor-bearing cell lines in vitro.* Jpn J Cancer Res, 1992. **83**(4): p. 373-8.
178. Armitage, R.J., et al., *Identification of a novel low-affinity receptor for human interleukin-7.* Blood, 1992. **79**(7): p. 1738-45.
179. Ferretti, E., et al., *Direct inhibition of human acute myeloid leukemia cell growth by IL-12.* Immunol Lett, 2010. **133**(2): p. 99-105.
180. Wang, L.N., et al., *Interleukin 21 and Its Receptor Play a Role in Proliferation, Migration and Invasion of Breast Cancer Cells.* Cancer Genomics Proteomics, 2015. **12**(5): p. 211-21.
181. Kondo, S., et al., *Human granulocyte colony-stimulating factor receptors in acute myelogenous leukemia.* Eur J Haematol, 1991. **46**(4): p. 223-30.
182. Bender, A.T., et al., *Differentiation of human monocytes in vitro with granulocyte-macrophage colony-stimulating factor and macrophage colony-stimulating factor produces distinct changes in cGMP phosphodiesterase expression.* Cell Signal, 2004. **16**(3): p. 365-74.
183. Lin, Z., et al., *Netrin-1 prevents the attachment of monocytes to endothelial cells via an anti-inflammatory effect.* Mol Immunol, 2018. **103**: p. 166-172.
184. Liu, H., et al., *Alternative splicing analysis in human monocytes and macrophages reveals MBNL1 as major regulator.* Nucleic Acids Res, 2018. **46**(12): p. 6069-6086.
185. Chen, R.F., et al., *Induction of IFNalpha or IL-12 depends on differentiation of THP-1 cells in dengue infections without and with antibody enhancement.* BMC Infect Dis, 2012. **12**: p. 340.

186. Kohro, T., et al., *A comparison of differences in the gene expression profiles of phorbol 12-myristate 13-acetate differentiated THP-1 cells and human monocyte-derived macrophage.* J Atheroscler Thromb, 2004. **11**(2): p. 88-97.
187. Tsuchiya, S., et al., *Induction of maturation in cultured human monocytic leukemia cells by a phorbol diester.* Cancer Res, 1982. **42**(4): p. 1530-6.
188. Lund, M.E., et al., *The choice of phorbol 12-myristate 13-acetate differentiation protocol influences the response of THP-1 macrophages to a pro-inflammatory stimulus.* J Immunol Methods, 2016. **430**: p. 64-70.
189. Hurtley, S.M., *Protein kinesis.* Science, 1996. **271**(5255): p. 1477.
190. Ou, W.J., et al., *Association of folding intermediates of glycoproteins with calnexin during protein maturation.* Nature, 1993. **364**(6440): p. 771-6.
191. Peterson, J.R., et al., *Transient, lectin-like association of calreticulin with folding intermediates of cellular and viral glycoproteins.* Mol Biol Cell, 1995. **6**(9): p. 1173-84.
192. Nauseef, W.M., S.J. McCormick, and R.A. Clark, *Calreticulin functions as a molecular chaperone in the biosynthesis of myeloperoxidase.* J Biol Chem, 1995. **270**(9): p. 4741-7.
193. Wang, C.H., et al., *Thrombospondin type I domain containing 7A (THSD7A) mediates endothelial cell migration and tube formation.* J Cell Physiol, 2010. **222**(3): p. 685-94.
194. Hou, Z., et al., *Expression, prognosis and functional role of Thsd7a in esophageal squamous cell carcinoma of Kazakh patients, Xinjiang.* Oncotarget, 2017. **8**(36): p. 60539-60557.
195. Usardi, A., et al., *The immunoglobulin-like superfamily member IGSF3 is a developmentally regulated protein that controls neuronal morphogenesis.* Dev Neurobiol, 2017. **77**(1): p. 75-92.
196. Synofzik, M., et al., *Absence of BiP co-chaperone DNAJC3 causes diabetes mellitus and multisystemic neurodegeneration.* Am J Hum Genet, 2014. **95**(6): p. 689-97.
197. Kulanuwat, S., et al., *DNAJC3 mutation in Thai familial type 2 diabetes mellitus.* Int J Mol Med, 2018. **42**(2): p. 1064-1073.
198. Agnello, V., et al., *Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor.* Proc Natl Acad Sci U S A, 1999. **96**(22): p. 12766-71.
199. Finkelshtein, D., et al., *LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus.* Proc Natl Acad Sci U S A, 2013. **110**(18): p. 7306-11.
200. Agrawal, P., et al., *A Systems Biology Approach Identifies FUT8 as a Driver of Melanoma Metastasis.* Cancer Cell, 2017. **31**(6): p. 804-819.e7.
201. Chung, C.T., S.L. Niemela, and R.H. Miller, *One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution.* Proc Natl Acad Sci U S A, 1989. **86**(7): p. 2172-5.
202. Ishihama, Y., J. Rappsilber, and M. Mann, *Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics.* J Proteome Res, 2006. **5**(4): p. 988-94.
203. Rappsilber, J., M. Mann, and Y. Ishihama, *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.* Nat Protoc, 2007. **2**(8): p. 1896-906.
204. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.

205. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat Methods, 2016. **13**(9): p. 731-40.
206. Chen, Y., et al., *Mascot-derived false positive peptide identifications revealed by manual analysis of tandem mass spectra.* J Proteome Res, 2009. **8**(6): p. 3141-7.
207. Pearson, T., et al., *A myeloma hybrid producing antibody specific for an allotypic determinant on "IgD-like" molecules of the mouse.* Eur J Immunol, 1977. **7**(10): p. 684-90.
208. Shulman, M., C.D. Wilde, and G. Kohler, *A better cell line for making hybridomas secreting specific antibodies.* Nature, 1978. **276**(5685): p. 269-70.
209. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 133-44.
210. Storoni, L.C., A.J. McCoy, and R.J. Read, *Likelihood-enhanced fast rotation functions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 3): p. 432-8.
211. Piechotta, A., et al., *Structural and functional analyses of pyroglutamate-amyloid-beta-specific antibodies as a basis for Alzheimer immunotherapy.* J Biol Chem, 2017. **292**(30): p. 12713-12724.
212. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
213. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.